REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	26-Sep-97	

4. TITLE AND SUBTITLE
EXTENSIONS TO REGRESSION ADJUSTMENT TECHNIQUES IN MULTIVARIATE STATISTICAL PROCESS MONITORING

5. FUNDING NUMBERS

6. AUTHOR(S)
Daryl James Hauck

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
ARIZONA STATE UNIVERSITY

8. PERFORMING ORGANIZATION REPORT NUMBER
97-025D

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA
2950 P STREET
WPAFB OH 45433

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A
Approved for public release
Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

19971006 073

14. SUBJECT TERMS	15. NUMBER OF PAGES
	145
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

EXTENSIONS TO REGRESSION ADJUSTMENT TECHNIQUES IN MULTIVARIATE STATISTICAL PROCESS MONITORING

by

Daryl James Hauck

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

ARIZONA STATE UNIVERSITY

December 1997

EXTENSIONS TO REGRESSION ADJUSTMENT TECHNIQUES

IN MULTIVARIATE STATISTICAL PROCESS MONITORING

by

Daryl James Hauck

has been approved

August 1996

APPROVED:

______________________________, Co-Chair

______________________________, Co-Chair

____

____

Supervisory Committee

ACCEPTED:

____

Department Chair

____

Dean, Graduate College

ABSTRACT

A common theme among the many existing multivariate statistical process monitoring (MSPM) methods is the recommendation that process knowledge be used to select a suitable monitoring procedure. Several methods possess the property of directional invariance, with shift detection performance depending only on the distance of a shift away from the target mean vector. This property is of special importance when characterizing a new process, or when available process knowledge suggests that shifts may occur in virtually any direction away from the target mean. In other cases, it is possible and may be desirable to increase a control scheme's sensitivity by using knowledge of the process structure and possible upset mechanisms to "aim" the control procedure. This research identifies a potentially common MSPM scenario and extends the idea of using process knowledge to determine an appropriate control statistic for assignable cause detection and identification.

Additionally, assumptions of normality and constant variance are imbedded in many statistical process monitoring procedures. For scenarios where monitoring with regression adjusted variables seems appropriate, but assumptions of normality and constant variance are violated, the use of prediction limits based on Generalized Linear Models theory was investigated and shown to be a potential improvement.

Finally, large capacity equipment may have several zones, making uniformity across zones an important objective. Furthermore, product delivery schedules may prevent machines from being dedicated to a single product specification, which adds

another source of variability to observed measurements. The use of regression adjustment with the addition of covariates to account for product specifications was investigated. For the three zone process investigated, shifts in one or two zones were shown to provide strong signals in the residuals. Furthermore, the residuals formed in this fashion were shown to provide a strong indication of uniformity dispersion effects across product specification levels.

DEDICATION

I would like to dedicate this work to my parents, Jim and Barbara Hauck, and to God who guided doctors in the successful early detection and removal of my father's cancer this spring. Gentlemen, know your PSA number.

ACKNOWLEDGEMENTS

The Air Force equivalent to a "Bull in a China Shop" is an officer who is "All Airspeed and No Heading." I am grateful to my committee co-chairs, Dr. Douglas Montgomery and Dr. George Runger, and to committee members Dr. Bert Keats and Dr. Dwayne Rollier for keeping me "on-course."

A special thanks to my wife, Lesley, and sons Nathan and Jayson, who provide the "Airspeed" and make every day exceptional.

I am thankful for and amazed by my fellow students whose creativity and diligence are an additional source of inspiration. In particular, I'd like to thank Mr. Dan McCarville for his assistance in obtaining industrial datasets.

A final note of gratitude to Dr. Roland Kankey and Major Kevin Grant of the Air Force Institute of Technology for providing this opportunity.

TABLE OF CONTENTS

		Page
LIST OF TABLES		x
LIST OF FIGURES		xi
CHAPTER		
1	INTRODUCTION	1
	Motivation for Multivariate Statistical Process Monitoring	1
	Importance of the Study	5
	Research Objectives / Organization	6
2	LITERATURE REVIEW	9
	Adaptations of "Classical" Univariate Methods	9
	Adaptations of Multivariate Analysis to Quality Control	25
	Other Methods	28
	Cross-Correlation vs. Auto-Correlation in SPC	32
	Summary	37
3	MULTIVARIATE STATISTICAL PROCESS MONITORING USING GROUPED REGRESSION ADJUSTED VARIABLES	38
	Introduction	38
	Background	39
	Grouped Regression Adjustment	44
	Example	46

CHAPTER Page

Other Impacts Associated With Grouping 52

Summary and Conclusions 53

Appendix 3.A. Proof of Independence 55

Appendix 3.B. Modification of Hawkins' Example 57

Appendix 3.C. Calculation of Non-Centrality Parameters . . . 60

Appendix 3.D. Computation of Non-Central Chi-Square Probabilities 64

Appendix 3.E. Principal Components Analysis of Example . . 65

Appendix 3.F. ARL Calculations for Level of Grouping . . . 67

4 MULTIVARIATE STATISTICAL PROCESS MONITORING INVOLVING NON-NORMAL DATA 69

Introduction 69

Background 70

Generalized Linear Models 74

Example 77

Autocorrelation 91

Other Considerations 91

Summary and Conclusions 92

Appendix 4.A. Supporting Data for Least-squares Model . . . 94

Appendix 4.B. Supporting Data for Least-squares Model using natural log transformation 97

Appendix 4.C. Supporting Data for Generalized Linear Model 102

CHAPTER Page

Appendix 4.D. Supporting Data for Generalized Linear Second-order Model 107

5 MONITORING UNIFORMITY ACROSS PROCESS ZONES AND CHANGING PRODUCT SPECIFICATIONS USING REGRESSION ADJUSTED VARIABLES 112

Introduction 112

Example 113

Other Considerations 119

Summary and Conclusions 120

Appendix 5.A. Zone Thickness Data 123

6 SUMMARY AND CONCLUSIONS 133

Contributions 133

Opportunities for Additional Inquiry 135

Conclusion 139

REFERENCES 140

LIST OF TABLES

Table		Page
1-1	Average Run Length of p Simultaneous Shewhart Charts	2
3-1	Principal Components of Cotton Fiber Data	49
3-2	Eigenvalue Analysis of Cotton Fiber Data Correlation Matrix	49
3-3	Comparison of χ^2 Chart on $\mathbf{U}_k$ vs. on $\mathbf{X}$ for composite shift case	52
3-4	Average Run Lengths of χ^2 Chart Combined Procedure Involving k Charts (Groups) with p_i Variables Each (p=12)	53
4-1	Summary Statistics for Fitted Models	83
4-2	Prediction Interval Length Comparison	85
5-1	Summary Statistics on Selected Process Observations	117

LIST OF FIGURES

Figure		Page
1-1.	Individual control charts for Y_1 and Y_2	3
1-2.	Joint plot of Y_2 vs. Y_1	4
2-1.	Control ellipse for two independent variables	33
2-2.	Control ellipse for two positively correlated variables	33
4-1.	Sheet resistivity vs. Layer thickness	78
4-2.	Probability plot of residuals from resistivity regressed on thickness	79
4-3.	Standardized residuals vs. predicted values	80
4-4.	Autocorrelation function of standardized residuals	80
4-5.	Probability plot of residuals from regression of ln(resistivity) on thickness	82
4-6.	Residuals versus predicted values from regression of ln(resistivity) on thickness	82
4-7.	95% prediction interval length vs. predicted resistivity	86
4-8.	95% prediction intervals from ln(resistivity) model	87
4-9.	95% prediction intervals from generalized linear model	88
4-10.	95% prediction limits for generalized linear model including second-order linear predictor	89
4-11.	99% prediction limits for generalized linear model including second-order linear predictor	90
4-12.	Run order plot of resistivity vs. 95% prediction limits	90

Figure		Page

5-1. Box and whisker plots of zone one residuals by recipe 115

5-2. Homogeneity of variance test for zone one residuals 115

5-3. Box and whisker plots of zone two residuals by recipe 116

5-4. Box and whisker plots of zone three residuals by recipe 116

5-5. Box and whisker plot of zone one residuals from model using "recipe-standardized" variables 121

5-6. Homogeneity of variance test for zone one residuals from model using "recipe-standardized" variables 121

CHAPTER 1

INTRODUCTION

Motivation for Multivariate Statistical Process Monitoring

There are many situations where it is important to simultaneously control many related quality characteristics (variables). These include (but are not limited to) examples of: inner and outer bearing diameter (Montgomery, 1991) where both determine the usefulness of the part; the cascade processes mentioned by Hawkins (1993) where the state of the j^{th} process step may effect the steps that come after; and the multiple stream processes mentioned by Mortell & Runger (1995) where the output of a filling head may be affected by causes specific to that head (clogging or wear) or by causes common to all heads on a machine (pressure or fluid viscosity changes).

Jackson (1985) provides a useful categorization of issues that differentiate multivariate techniques from independently using "classical" univariate techniques on each variable:

> 1) They will produce a single answer to the question: "Is the process in control?"; 2) The specified type I error will be maintained; and 3) These techniques will take into account the relationship between these two [or more] variables.

The point made in 1) primarily refers to the avoidance of effort involved in maintaining many control charts simultaneously. The point in 2) refers to a typical problem involved in simultaneous inference -- if we have two control procedures, each with Type I error probability $\alpha = .05$, this level is not maintained for a simultaneous inference about the state of both. Montgomery (1991) gives the Type I error for the joint control procedure as

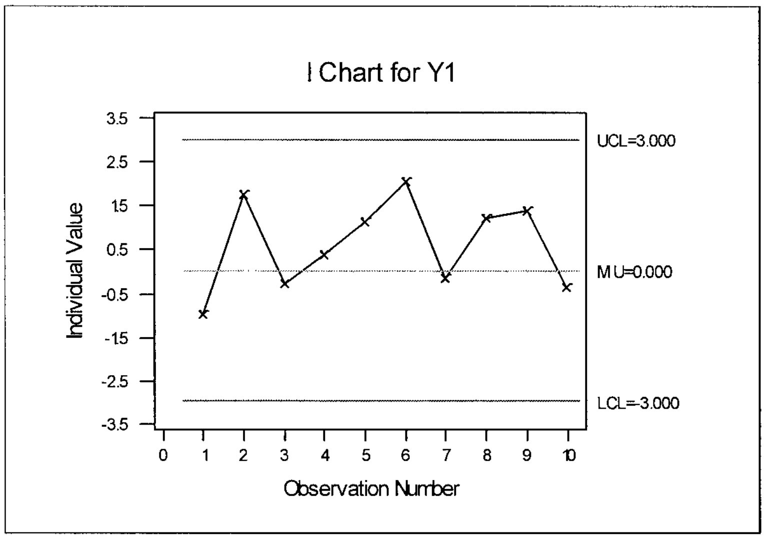
$\alpha' = 1-(1-\alpha)^p$ where p = the number of independent quality characteristics. He also states that if the p quality characteristics are not independent, then there is no way to measure the distortion of the joint procedure. This distortion also translates as shorter in-control average run lengths (higher false alarms). Table 1-1 shows Runger's (1996a) example of the significance of this distortion in typical Shewhart control charts using 3-sigma limits run simultaneously. Multivariate techniques are designed to avoid this distortion.

Perhaps the strongest benefit of multivariate methods lies in 3), the ability to exploit relationships between variables. While the problem with simultaneous univariate methods in 2) results in degradation of performance, it is possible that ignoring relationships between variables will miss unusual process states altogether. Figure 1-1 (Runger, 1996a) demonstrates the case of two univariate charts showing both variables are within control limits. Figure 1-2 (Runger, 1996a) uses the same data considered jointly to demonstrate that one data record is violating the usual relationship between the two variables and should be investigated.

Up to this point, only advantages of multivariate methods have been stated to motivate their study. Hawkins (1991, 1993) points out that in spite of their statistical power many multivariate techniques do not indicate the cause of a signal. This is a

Table 1-1. Average Run Length of p simultaneous 3-sigma Shewhart charts (Runger, 1996)

Number of Variables [p]	Data Records Between False Alarms
1	370
20	20
200	2



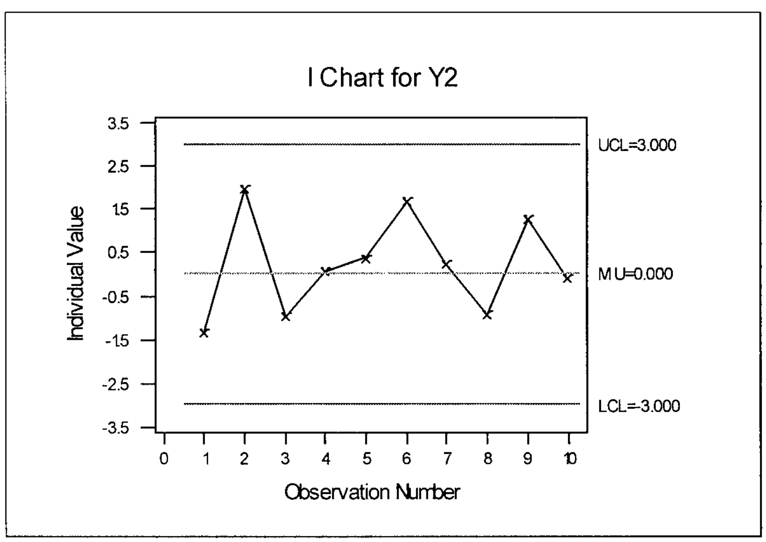


Figure 1-1. Individual control charts for Y_1 and Y_2. (Runger, 1996a)

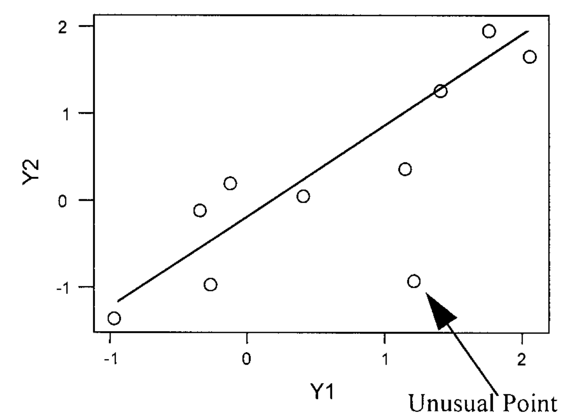


Figure 1-2. Joint plot of Y_2 vs. Y_1. (Runger, 1996a)

by-product of reducing the dimensionality of the problem to single (or a few) indicators. This aspect of finding the problem must also be considered.

While relationships (correlations) between variables are advantageous for multivariate control schemes, autocorrelation within variables can hurt statistical process control schemes (Montgomery and Mastrangelo, 1991; Harris and Ross, 1991), causing a dramatic increase in false alarms. Sensitivity to autocorrelation is another important factor for consideration.

In summary, it is desirable to condense information so that the process is more easily monitored; retain enough information so that assignable cause may be determined when the control scheme does signal; achieve a suitable balance between high in-control average run lengths (low false alarm rate) and fast detection of real shifts in some variable or combination of variables; maintain specified Type I error levels; be sensitive to shifts

in all directions (unless process knowledge suggests structured shifts); and be relatively insensitive to autocorrelation. This is a tall order, with several of the objectives in conflict; however, these desirable properties serve as useful evaluation criteria for contrasting performance of proposed methods.

Importance of the Study

While Hotelling's (1947) bombsight data analysis is often attributed as the first work in the multivariate quality control area, it is not an "old" field by any means. Handling more than just a few variables is computationally intensive, and interest in multivariate quality control techniques waned until the 1980's when the availability of computers and a renewed interest in quality became more widespread [of 37 papers referenced by the Lowry and Montgomery (1995) survey paper, only three are prior to 1984]. Both the need and opportunity for original contribution in this field are high.

An emerging theme among several of the existing multivariate statistical process monitoring (MSPM) methods is the recommendation that process knowledge should be used to select a suitable monitoring procedure. The first consideration involves considering any unique structure in the process itself. For example, Mortell and Runger (1995) propose a model and control strategies for multiple stream processes to detect shifts in the mean of all streams vs. a shift in a single stream of a multi-head filling machine. Hawkins (1993) modified his original regression adjustment approach to be more suitable for monitoring cascade (sequential value added) processes.

Other considerations involve expectations concerning the nature of a potential process upset: 1) whether a shift is likely to occur in one, several, or all of the quality characteristics being monitored; 2) whether the relationships between quality characteristics are maintained under shift conditions (model-fixed) or changed (model-void) (Runger, 1996b); and 3) whether a shift is likely to occur in any direction within the region of interest or towards a known out-of-control state (Healy, 1987). The possible combinations of these factors mounts quickly, and no single monitoring method emerges as an overall superior approach.

Research Objectives / Organization

A literature review is the necessary first step in determining research opportunities, assessing their importance, and becoming "schooled" in domain-specific knowledge and terminology. While the focus in the Chapter 2 literature review is to emphasize material related to specific research objectives mentioned below, additional material uncovered in the course of this review is also presented to provide an interested reader with additional references in the multivariate process monitoring arena.

In processes where model-void shifts are expected, regression adjustment techniques proposed by Mandel (1969) and Hawkins (1991, 1993) have shown promise in improved sensitivity and diagnosis of assignable cause. These methods rely on the use of process knowledge and the specific process structure. Necessarily limiting assumptions are required in their formulation. An objective of this research is to remove some of these limitations. Chapter 3 identifies a potentially common multivariate process

monitoring scenario and extends the idea of using process knowledge to determine an appropriate control statistic for assignable cause detection and identification. The result is a generalized regression adjustment procedure applicable to a wider class of sequential-value-added (cascade) processes.

Regression adjustment techniques based on least-squares estimators depend on the assumption that the quality characteristics of interest are normally distributed. Furthermore, other methods (T^2, MCUSUM, MEWMA) either depend on p-variate normality assumptions or, as a minimum, have used normally distributed variates in calcualtions/simulations of shift detection performance as measured by the Average Run Length (ARL). An objective of this research is to create a method for handling non-normal data in situations that otherwise seem well-suited to regression adjustment procedures. Chapter 4 describes a semiconductor process where one variable (thickness) is controlled directly by the process and is normally distributed; and a second variable, resistivity, is usually a function of thickness but follows a skewed distribution and may be moved in a "model-void" fashion from its target by causes unrelated to thickness, such as contamination. The use of Generalized Linear Models to appropriately fit the relationship between thickness and resistivity in spite of the latter's non-normality and non-constant variance is explored.

In some manufacturing processes, large capacity equipment may have several zones, with slight differences between zones adding an additional source of variability. An important objective is ensuring that uniformity between zones is maintained to the

highest possible extent. Furthermore, due to product delivery schedules, it may not be feasible to dedicate a single machine to a particular product formulation. Process attributes may need to be adjusted from run to run to meet target product specifications. Chapter 5 explores monitoring uniformity with regression adjusted variables, with modification to the regression models to adjust for varying product specification.

Finally, Chapter 6 summarizes the research findings, and presents ideas that may lead to additional inquiry in the multivariate statistical process monitoring field.

CHAPTER 2

LITERATURE REVIEW

Multivariate control methods proposed to date may be classified into three broad categories: 1) Multivariate extensions of "classical" univariate SPC methods -- T^2, MCUSUM, MEWMA, Regression Adjustment, etc.; 2) Adaptations of multivariate analysis techniques to the quality control application -- Principal Components, Partial Least Squares, Discriminant Analysis, etc.; and 3) Other --Graphical Techniques, Neural Nets, Pattern Recognition, etc. After current methods are reviewed, a section emphasizing the differences between and the impacts of cross-correlation and autocorrelation is presented in order to further distinguish between them since cross-correlation is not considered in univariate SPC.

Adaptations of "Classical" Univariate Methods

Hotelling's T^2

In 1931, Harold Hotelling published "The Generalization of Student's Ratio," for the testing of hypotheses about the location of the means of multivariate distributions (Hotelling, 1947). Jackson (1959), restated Hotelling's method in the matrix form that is more commonly used today. The following summary is from Lowry and Montgomery (1995). Assuming that p quality characteristics are jointly distributed as p-variate normal and that random samples of size n are collected, Hotelling's chart signals a statistically significant shift in the mean when

$$\chi_i^2 = (\mathbf{X}_i - \mu_0)' \Sigma^{-1} (\mathbf{X}_i - \mu_0) > h_1 \qquad (2\text{-}1)$$

where h_l is the selected control limit. This procedure is considered a natural multivariate extension of the univariate Shewhart chart. Most often, Σ and μ_0 are unknown and must be estimated by $\overline{\mathbf{X}}$ and $\mathbf{S}$ using historical data {when these are substituted into (2-1), the resulting expression is often referred to as Hotelling's T^2}. When μ and Σ are known [or estimated from a relatively large number of preliminary samples] the upper control limit for the χ^2 chart [T^2] is often set to UCL = $\chi^2_{\alpha,p}$. When this is not the case, Hotelling showed that T^2 is related to the F-distribution by

$$T^2_{p,n,\alpha} = \frac{p(n-1)}{n-p} F_{p,n-p,\alpha} \tag{2-2}$$

The average run length (ARL) properties of the χ^2 chart depend only on the distance of the shift from the target mean vector, μ_0. This distance is given by

$$\lambda_\mu = \sqrt{(\mu - \mu_0)' \Sigma^{-1} (\mu - \mu_0)} \tag{2-3}$$

Under shift conditions, the χ^2 statistic follows the non-central chi-square distribution with the non-centrality parameter equal to λ_μ^2. Lowry and Montgomery (1995) and Wierda (1994) give additional guidance for adjustments to the sample statistic and reference distribution depending on the control phase (Phase I - retrospective test for control initiation versus Phase II - continued monitoring) and the sample size (a single observation on each characteristic, or the sample mean of several observations on each characteristic).

Hawkins (1991) mentions that "although the T^2 is the optimal single-test statistic for a general multivariate shift in the mean vector, it is not optimal for more structured mean shifts - for example shifts in only some of the variables." Structured shifts are common, and this motivated Hawkins development of regression adjustment techniques described later. Since the T^2 combines all the information into a single statistic, it does not give insight into the cause(s) of the indicated shift. A common approach has been to retain all individual charts for diagnosis. This approach has its own set of shortcomings and will be discussed in the "graphical techniques" section.

Lowry and Montgomery (1995) point out that because the Hotelling T^2 is based on only the most recent observation, it is insensitive to small and moderate shifts in the mean vector. Multivariate CUSUMs and EWMAs contain information on previous periods and are more sensitive to small shifts.

Multivariate Cumulative Sum (MCUSUM)

Several authors propose multivariate CUSUM procedures. Woodall and Ncube (1985) consider the use of several univariate CUSUM procedures to be a single multivariate CUSUM procedure. The focus of their paper is on an analytical technique for computing the Average Run Length of such an approach (with the run length determined by the first chart to signal) under assumptions of independence between variables. For dependence in the original variables, they recommend transformation to principal component analysis to obtain independent components (see the section on "Adaptations of Multivariate Analysis to Quality Control" for a description of Principal

Components Analysis). For $p = 2$, the use of independent CUSUMs on the principal components is shown to detect small shifts more quickly than the T^2 chart. Assuming the original variables are correlated, the approach based on principal components analysis would pick up "relationship" shifts that the approach on original variables would miss.

From an interpretability perspective, several authors [Lowry and Montgomery, (1995); Hawkins, (1991) to name a few] point out that principal components may not always have an interpretable meaning with respect to the process (especially as p gets large). While the approach on original variables would be easy to interpret when variables are outside individual control limits, this approach would miss changes in relationships between variables that occurred inside the control limits on each variable. Woodall and Ncube dismiss this problem saying "in many cases a shift in correlation coefficient has no adverse effect on quality and does not require any corrective action to be taken." Pignatiello and Runger (1990) note that it may not be possible to provide corrective action on a single variable without affecting one or more of the other variables, and offer the opposing viewpoint that relationships between all variables must be considered to properly interpret a signal and determine corrective action.

To summarize Woodall and Ncube from an "interpretability" standpoint, shifts that are outside univariate control limits would be clearly seen. Changes in relationships within individual univariate control limits would be missed. Using principal components in the separate CUSUMs would allow detection of relationship changes within individual univariate control limits, but may complicate determination of assignable cause.

Healy's (1987) conversion from a Multivariate CUSUM to a univariate CUSUM (when a shift is expected in only one direction defined by μ_B) is very useful in that it allows existing univariate CUSUM theory to be used in selecting the control limit, H, and the initial CUSUM value, S_0 [important for fast initial response properties], to achieve desired Average Run Length (ARL) properties. Healy's CUSUM for detecting a shift in the multivariate normal mean is

$$S_n = \max\,(S_{n-1} + \mathbf{a}'\,(\mathbf{x}_n - \mu_G) - .5D, 0) > H \tag{2-4}$$

where

$$\mathbf{a}' = (\mu_B - \mu_G)'\,\Sigma^{-1} / [(\mu_B - \mu_G)'\,\Sigma^{-1}(\mu_B - \mu_G)]^{1/2} \tag{2-5}$$

and

$$D = [(\mu_B - \mu_G)'\,\Sigma^{-1}(\mu_B - \mu_G)]^{1/2} \tag{2-6}$$

where "D" is the "noncentrality parameter." D measures the shift in the multivariate mean in terms of a statistical distance (if the variables are uncorrelated, D becomes a "standardized" Euclidean distance). Healy showed that the ability of this procedure to discriminate between μ_G and μ_B depends on μ_G , μ_B and Σ only through D -- it does not depend on the number of variables, p. He states the reason is that this procedure only looks at one particular direction. It should be clear that the shift doesn't have to be only in one variable -- the shift is in a multivariate mean from μ_G to μ_B with distance D.

Healy's procedure requires an estimate of the mean in the "bad" state -- in his example, product knowledge suggested this value directly. When this is not known, he

suggests using Crosier's MCUSUM (referencing an unpublished 1986 paper which became part of Crosier's 1988 Technometrics paper described below).

Healy stated "My procedure makes sense when shifts in only a few known directions are to be expected." If shifts in other directions are likely, Healy recommends adding CUSUMs of linear combinations that are orthogonal. If particular process "failure modes" continually lead to a similar upset, then implementing this procedure makes a lot of sense. However, as the number of process steps involved increases it seems reasonable that the potential for several different failure modes is also increased, with interaction between them influencing the size of any particular upset.

Healy suggests adding another one-sided CUSUM to form an ordinary two-sided CUSUM procedure where the elements of μ_B for the second CUSUM are based on the size of the undesirable shift in the opposite direction as before. For shifts in other directions, additional CUSUMs are added in such a manner that the linear combinations in $\mathbf{a}'\mathbf{x}$ are orthogonal. For p = 2, this would amount to four individual CUSUMs, with the number rising to eight for p=3 -- herein lies an important disadvantage. Maintaining this many separate CUSUMs raises the false alarm rate of the overall procedure, unless each CUSUM is "de-sensitized" so that the overall Type I error rate is maintained. Changing the control limit to reduce false alarms hinders shift detection performance.

From an interpretation viewpoint, a signal in Healy's approach would indicate a shift similar in structure to that specified for μ_B. "Similar" because a large enough shift in directions at some angle (less than "orthogonal") to μ_B could also cause a signal. If

additional CUSUMs are added such that $\mathbf{a}'\mathbf{x}$ are orthogonal, and you cannot add any others that would be orthogonal, then a signal would be narrowed down "directionally," but there are many ways for different sizes and combinations of shifts to cause the signal. Conceptually, univariate control charts within limits with a CUSUM on the linear combination signaling may indicate a relationship change; however, since Healy's approach determined the linear combinations by specification of μ_B rather than on historical data, this reasoning may not hold.

Crosier (1988) considers two approaches to multivariate CUSUMs. The first is referred to as COT (the CUSUM of the scalars T_n, which are the positive square roots of the familiar T^2 statistic). This CUSUM is given by:

$$S_n = \max(0, S_{(n-1)} + T_n - k) \tag{2-7}$$

This scheme signals when $S_n > h$.

The second form is created by replacing scalar quantities of a univariate scheme with the vectors such that:

$$\mathbf{s}_n = \max\,(\mathbf{0}, \mathbf{s}_{n-1} + (\mathbf{x}_n - \mathbf{a}) - \mathbf{k}) \tag{2-8}$$

where $\mathbf{a}$ contains the target mean for each variable. Crosier notes problems associated with selecting $\mathbf{k}$ and how to interpret taking a maximum of a vector and a null vector. $\mathbf{k}$ needs to be determined so that it "shrinks" $\mathbf{s}_{n-1} + (\mathbf{x}_n - \mathbf{a})$ towards zero. Crosier determined a suitable $\mathbf{k}$ is:

$$\mathbf{k} = (k/C_n)(\mathbf{s}_{n-1} + \mathbf{x}_n - \mathbf{a}) \tag{2-9}$$

where C_n is the length of $(\mathbf{s}_{n-1} + \mathbf{x}_n - \mathbf{a})$ given by:

$$C_n = [(\mathbf{s}_{n-1} + \mathbf{x}_n - \mathbf{a})' \, \Sigma^{-1}(\mathbf{s}_{n-1} + \mathbf{x}_n - \mathbf{a})]^{1/2} \tag{2-10}$$

To interpret picking the maximum of a vector with respect to a null vector Crosier defined:

$$\begin{aligned} \mathbf{s}_n &= \mathbf{0} && \text{if } C_n \leq k \\ \mathbf{s}_n &= (\mathbf{s}_{n-1} + \mathbf{x}_n - \mathbf{a})(1 - k/C_n) && \text{if } C_n > k \end{aligned} \tag{2-11}$$

The multivariate control scheme signals when $Y_n > h$, where:

$$Y_n = [\mathbf{s}_n' \, \Sigma^{-1}\mathbf{s}_n]^{1/2} \tag{2-12}$$

Crosier's paper focused on ARL performance comparisons rather than signal interpretation. Regarding interpretation, Crosier states "the CUSUM vector of a multivariate CUSUM scheme could be examined to determine the nature of the process problem." Y_n is excluded from interpretation as it is a scalar composite indicator that cannot assign cause. Observing each of the elements of $\mathbf{s}_n$ will only help when variables are outside of their univariate control limits. A signal caused by a change in the relationship between variables (with each variable remaining within its own limits) would not show in the univariate components of $\mathbf{s}_n$, and could be interpreted as a false alarm.

The interpretation problem for COT is even more difficult. It is a CUSUM based on the positive square root of T^2, which is itself a composite indicator. Furthermore, Hawkins (1991) states "Measures based on quadratic forms (like T^2) also confound mean shifts with variance shifts and require quite extensive analysis following a signal to determine the nature of the shift." A T^2 decomposition approach proposed by Mason, Tracy and Young (1995) does quite well at interpreting the troublesome variables, but

information relative to location versus variability shifts is not highlighted in their procedure. An improved algorithm requiring fewer computations is in press (Mason *et. al.*, 1997). For cascade processes, downstream variables are also affected by causes at an upstream change. Since the downstream means also change, diagnostic methods based on T^2 decompositions are likely to assign cause in more places than it originated.

Pignatiello and Runger also consider two formulations for multivariate CUSUMs. The first is based on accumulating differences between the sample average and the target value for the mean, then applying a quadratic distance measure; whereas the second method calculates a quadratic distance for each **X**, then accumulates the quadratic form.

The first method is called MC1 which takes the form:

$$MC1_t = \max \{ \|\mathbf{C}_t\| - kn_t, 0\} \tag{2-13}$$

where

$$\begin{aligned} n_t &= n_{t-1} + 1 && \text{if } MC1_{t-1} > 0 \\ &= 1 && \text{if otherwise,} \end{aligned} \tag{2-14}$$

$$\| \mathbf{C}_t \| = (\mathbf{C}_t' \, \Sigma^{-1} \, \mathbf{C}_t)^{1/2}, \tag{2-15}$$

$$\text{and } \mathbf{C}_t = \sum_{i=t-n_t+1}^{t} (\mathbf{X}_i - \mu_0). \tag{2-16}$$

Though formulated differently, the ARL performance of MC1 is similar to Crosier's Y_n.

The second method is referred to as MC2 and takes the form:

$$MC2_t = \max \{0, MC2_{t-1} + D_t^2 - k\} \tag{2-17}$$

where

$$D_t^2 = (\mathbf{X}_t - \mu_0)' \Sigma^{-1} (\mathbf{X}_t - \mu_0). \tag{2-18}$$

Pignatiello and Runger noted that MC2 has ARL performance similar to Crosier's COT, even though Crosier's formulation involves the positive square root of T^2 (note: D^2 and T^2 are essentially the same quantity).

From an interpretability standpoint, MC2 suffers from the same difficulty as COT -- they are based on composite indicators that do not directly tell us anything about the individual variables. As previously mentioned, these measures also confound location and variability shifts.

Pignatiello and Runger recommend MC1 due to its better ARL performance. After an illuminating discussion on the need for considering relationships between variables in interpreting shifts and deciding upon corrective action, the suggestion for interpretation is supplementing the MC1 chart (for monitoring) with multiple univariate control charts (for interpretation). As previously discussed, univariate charts will not indicate relationship changes that can occur while observations remain within univariate control limits.

The potential problem of CUSUM inertia is described along with that of MEWMAs below.

Multivariate Exponentially-Weighted Moving Average (MEWMA)

Lowry *et al.* (1992) propose a multivariate extension of the univariate EWMA. The MEWMA consists of vectors of EWMAs:

$$\mathbf{Z}_i = \mathbf{R}\mathbf{X}_i + (\mathbf{I} - \mathbf{R})\mathbf{Z}_{i-1} \tag{2-19}$$

where $\mathbf{Z}_o = \mathbf{0}$ and $\mathbf{R} = \text{diag}(r_1, r_2, \ldots, r_p)$ [smoothing constants], $0 < r_j < 1$ (r_j are generally set to be equal), j = 1, 2, ..., p. These vectors are used to form T^2 statistics in the form:

$$T_i^2 = \mathbf{Z}_i' \Sigma_{Z_i}^{-1} \mathbf{Z}_i \qquad (2\text{-}20)$$

These statistics are compared to tabular reference values selected to achieve desired ARL properties. A useful result is that when the MEWMA chart signals, the $\mathbf{Z}_i$ vectors give some indication of the direction. Lowry *et al.* show that MEWMA ARLs are better than the Hotelling T^2 and MCUSUM in detecting an initial out of control condition, and are as good as MCUSUM methods when the process is initially in-control but shifts later.

Both the MEWMA and the MCUSUM suffer from inertia -- if the process was operating near one control limit, then began a shift towards the other, these procedures would be slow to react. This occurs because the statistics include weighted information on prior periods. Lowry *et al.* recommend a Hotelling T^2 be used concurrently to protect against inertia.

With the exception of Healy's (1987) method, the procedures mentioned thus far have the important property of directional invariance; that is, the average run length (ARL) to detect a shift depends only on the distance of the shift from the multivariate mean (this is true for the MEWMA only as long as the weights used in each univariate EWMA are equal). This directional invariance property is of special importance when characterizing a new process, or when available process knowledge suggests that shifts may occur in virtually any direction away from the target mean. In other cases, it is possible and may be desirable to increase a control scheme's sensitivity by using

knowledge of possible upset mechanisms to "aim" or sensitize the control procedure. Intuitively speaking, why expend resources monitoring regions towards which the process mean is unlikely to shift? Several techniques are available that use process knowledge to sensitize shift detection and/or improve diagnosing assignable cause.

Regression Control Charts

Mandel (1969) introduced the idea of a regression control chart to adjust the quality characteristic of interest (post office man-hours expended) to account for an outside covariate (mail volume) -- parallel control limits were based on a constant times residual mean squared error (versus using the regression confidence limits). In this scenario, the man-hours mean is variable, and it is only important to identify offices or changes at a specific office that violate the "usual" relationship between man-hours and mail volume -- "model fixed" changes in mean man-hours were expected and not considered important by management. Note that the property of directional invariance for the control mechanism is deliberately eliminated.

Hawkins (1991) synthesized Mandel's regression adjustment, with Healy's (1987) approach for maximizing sensitivity to a structured shift, using the additional assumption that a shift would occur in only a single variable. The procedure involves a control statistic for each variable that is the "residual when X_i is regressed on all other components of **X** rescaled to unit variance:"

$$Z_i = \frac{\left[(X_i - \mu_i) - \sum_{j \neq i} \beta_{ij}(X_j - \mu_j)\right]}{\sqrt{\tau_{ii}}} \tag{2-21}$$

where τ_{ii} is the variance of the residuals given by

$$\tau_{ii} = \sigma_{ii} - \sum_{j \neq i} \beta_{ij} \sigma_{ij} \tag{2-22}$$

and σ_{ij} is the i,jth element of the initial variance-covariance matrix, Σ_0. The Z_i are used to detect location shifts in the ith variable.

Hawkins showed an important relationship between T^2, **X** and **Z** is:

$$T^2 = \sum_{i=1}^{p} W_i, \quad \text{where } W_i = (X_i - \mu_i) Z_i \tau_{ii}^{-1/2} \tag{2-23}$$

Hawkins recommends basing control charts for variability shifts on W_i, noting that they represent the scaled product of the deviation of variable i on the **X** scale and its deviation on the **Z** scale.

CUSUM control for location on each Z_i (and W_i) may be performed individually or as a group. For individual charts:

$$\begin{aligned} L^{+}_{in} &= \max(0, L^{+}_{i,n-1} + Z_{ni} - k) \\ L^{-}_{in} &= \min(0, L^{-}_{i,n-1} + Z_{ni} + k) \end{aligned} \tag{2-24}$$

$$\begin{aligned} S^{+}_{in} &= \min(0, S^{+}_{i,n-1} + W_{ni} - k) \\ S^{-}_{in} &= \min(0, S^{-}_{i,n-1} + W_{ni} + k) \end{aligned} \tag{2-25}$$

It should be clear that a control chart on Z_i is not a univariate chart on variable i:

> Each Z_i is a linear combination of all of the measures of **X**, in general all with non-zero coefficients. This means that a shift in any component of **X** can lead to a displacement in all of the components of **Z** (Hawkins, 1993).

{note: the above quote applies to non-cascade process -- for cascade processes he recommends using Y_j, which is defined as a linear combination of X_1 through X_{j-1}.}

For group control Hawkins proposes:

$$MCZ = \max\{\max(L^+_{ni} - L^-_{ni})\} \tag{2-26}$$

and

$$ZNO = \sum_{i=1}^{p}(L^+_{ni} + L^-_{ni})^2 \tag{2-27}$$

(note: group control can be performed on S^+ and S^- charts as well). Hawkins procedures had the best average ranking in ARL performance of the methods considered in his paper, with ZNO being the best (methods compared included Woodall and Ncube's multiple univariate scheme, and both of Crosier's approaches {roughly equivalent in performance to Pignatiello and Runger's approaches}). Situations that degraded Z-based ARL performance were: shifts in the same direction for pairs of highly correlated variables, and shifts proportional to the two leading principal components. Hawkins' procedures ranked near the last in only five of the 30 cases examined. MCZ ranked first five times, and ZNO ranked first 15 times.

In a fashion similar to that recommended by Pignatiello and Runger, Hawkins suggests using ZNO for monitoring, and the individual Z_i for interpretation. Relying on Z_i for interpretation does not sacrifice relationship information since they are linear combinations of the X_i. Hawkins shows that a shift of δ in the mean of X_i will result in a shift of $\delta(1- R_i^2)^{-1/2}$ in the Z_i (where R_i^2 is the multiple correlation between X_i and all

other components of X). As the correlation increases, the shift in X_i is magnified in the Z_i.

From an interpretability viewpoint Hawkins states "the vector Z also has the valuable interpretive property that signals given are for shifts in the mean, or shifts in the variance, of particular variables rather than global signals indicating some unspecified departure from control" {note: the indications for variance shifts actually come from **W**}. This was demonstrated in an example for five variables, where a variance shift was induced for variable 1, and a location shift was introduced into variable 5. Charts for Z_i and W_i associated with variables 2 - 4 remained within control limits, while charts W_1 and Z_5 correctly signaled.

Hawkins (1993) modified his technique for application in sequential-added-value (cascade) processes by using only "upstream" observations as the independent variables in the regression adjustment.

In contrast to Hawkins' use of simultaneous univariate CUSUMs, Timm (1996) applies "stepdown" finite intersection tests (FIT) to the same regression adjusted variables. Timm reparameterizes the hypotheses that all means are equal to their target value into the hypothesis:

$$H = \bigcap_{i=1}^{p} H_i^* \text{ where } H_i^* : \eta_i = \eta_{0i} \tag{2-28}$$

and $\eta_{i+1} = \mu_{i+1} - \beta\mu_i$. It's easy to see that η_{i+1} is the residual obtain by subtracting the predicted value of a mean (based on the observed means from previous steps) from it's observed value at the current step. The test statistic for testing (2-28) is:

$$F_i = \frac{(\hat{\eta}_i - \eta_{0i})^2(n-i)}{d_i s_i^2} \tag{2-29}$$

where $d_i s_i^2$ is the estimated variance of $\hat{\eta}_i$. The hypothesis in (2-28) is rejected if

$F_i > f_{i\alpha,1,n-i}$, where α is chosen so that $\prod_{i=1}^{p} \Pr(F_i \le f_{i\alpha,1,n-i}|H) = 1-\alpha$. For cascade process, an a priori order exists and the step-down Fit is performed one time. Otherwise, when an a priori order is unknown, step down FITs are performed on all possible orderings and type I error for each FIT must be set to control the familywise error rate. For a cascade process, cause is assigned to the step that first rejects an H_i. Without a known a priori order, all orderings would need to be completed. Timm states this procedure is optimal when an a priori order is known.

Zhang's (1984, 1985) cause selecting control charts incorporate a two-step approach that monitors incoming quality and also monitors outgoing quality adjusted for the incoming quality. The adjustment may be made by theoretical or empirical models. The combinations of possible signals in each chart help to diagnose whether incoming, outgoing, or both measures contain assignable cause. Wade & Woodall (1993) showed that using regression prediction limits instead of confidence limits provided more predictable ARL performance in this procedure.

Adaptations of Multivariate Analysis to Quality Control

Principal Components Analysis (PCA)

Principal components (Hotelling, 1933) decomposes data into orthogonal linear combinations based on the eigenvectors of the ordered (greatest to least) eigenvalues that describe the present variability in the data. The eigenvalues of the covariance matrix are obtained by solving the "characteristic" equation (Jackson, 1980):

$$|\,\mathbf{S} - \lambda\mathbf{I}\,| = 0 \tag{2-30}$$

Eigenvectors associated with each eigenvalue are obtained by solving the equations:

$$\begin{gathered}[\mathbf{S} - \lambda_i\,\mathbf{I}]\mathbf{t}_i = 0 \\ \mathbf{u}_i = \mathbf{t}_i \,/\, (\mathbf{t}_i'\,\mathbf{t}_i)^{1/2}\end{gathered} \tag{2-31}$$

Many software packages contain eigen-analysis routines, making the calculations of principal components from the covariance matrix very straightforward. The number of principal components available is equal to the number of variables. Monitoring methods based on a full set of principal components retain the property of directional invariance. Jackson (1980) notes that most process variability is often captured in the first few principal components, and that the task of monitoring a large number of variables can be reduced into monitoring a few principal components. PCA has the advantage of attaining independence between variables and reducing dimensionality, but sometimes these components don't have an interpretable meaning (Lowry and Montgomery, 1995).

Scranton *et. al.* (1996) showed the improvement in ARL performance that can be obtained by using an MEWMA on a subset of principal components.

Partial Least Squares (PLS)

Partial Least Squares regression (sometimes called projection to latent structures) was developed in the late sixties by H. Wold for use in econometrics, and was initially applied to chemical engineering by Kowalski, Gerlach and H. Wold, with further development by S. Wold and Martens (Geladi and Kowalski, 1986). The method creates linear combinations of explanatory variables (**X**) that are predictive of an external data set containing other variables of interest (**Y**). An iterative estimation technique essentially reconciles (by increasing correlation between) factor loadings that independently explain the most variability in the separate coordinate systems (Stahle and S. Wold, 1988). The results are models (Garthwaite, 1994):

$$\mathbf{Y}_k = \beta_{k0} + \beta_{k1}T_1 + \ldots + \beta_{kp}T_p \tag{2-32}$$

where the components $T_1, \ldots, T_p$ are linear combinations of the X variables. The same components occur in the model for each Y variable, only the regression coefficients are different for each Y. PLS is very similar to PCA, but the plane in **X** is "tilted" to be more predictive of the variables in **Y** -- while the variability in X is considered (and may be used to identify outliers prior to prediction) it no longer dominates factor scores. If a variable does not exhibit as much variability but is highly predictive of Y, it may receive a stronger loading. Additional detail on the concept, estimation algorithmn, and comparisons to other multivariate analysis methods may be found in Stahle and S. Wold (1988), Garthwaite (1994), Kresta, *et. al.* (1991), and Geladi and Kowalski (1986).

Garthwaite's (1994) simulation study suggests PLS is most useful when the number of explanatory variables and the error variance are large.

This method provides an advantage of monitoring a reduced number of components, but diagnosis of assignable cause may be complicated as a signal is caused by a high "score" in a linear combination containing factor loadings that are themselves linear combinations of other factors.

Discriminant Analysis

Murphy (1987) proposes using Hotelling's T^2, and recommends applying discriminant analysis after an "out of control" signal to determine a subset p_1 of the original p variables that are suspected of causing the signal. Given that a vector of observations $(\bar{x}^*)$ signals out of control $[T^2(\bar{x}^*) > K]$, $\bar{x}^*$ is partitioned into $(\bar{x}^{*(1)}, \bar{x}^{*(2)})$ where $\bar{x}^{*(1)}$ is the p_1 subset of the p variables which are suspected of causing the signal and $\bar{x}^{*(2)}$ is the remaining p_2 variables. The full squared distance is measured by

$$T_p^2 = T^2(\bar{x}^*) = n(\mu_0 - \bar{x}^*)'\Sigma^{-1}(\mu_0 - \bar{x}^*) \qquad (2\text{-}33)$$

and the reduced squared distance corresponding to the p_1 subset is

$$T_{p1}^2 = T^2(\bar{x}^{*(1)}) = n(\mu_0^{(1)} - \bar{x}^{*(1)})'\Sigma_{(1)}^{-1}(\mu_0^{(1)} - \bar{x}^{*(1)}) \qquad (2\text{-}34)$$

where μ_0 and Σ are partitioned as is $\bar{x}^*$. If the Difference ($D = T_p^2 - T_{p1}^2$) is large, the null hypothesis that the p_1 subset caused the signal is rejected. This is equivalent to testing that the p_1 subset "discriminates" just as well as the full set of p variables.

Murphy showed that when H_o is true $D \sim \chi^2_{p2}$. If (μ, Σ) are estimated, Murphy references Seber (1984) for the appropriate F-test. A disadvantage of this procedure is that it relies on the T^2 which is known to signal slower than MCUSUMs or MEWMAs.

Other Methods

Graphical Methods

The high dimensionality of multivariate quality control generally makes plots of joint regions impossible. One-at-a-time charts can maintain the independent status of each variable, and Blazek *et al.* (1987) describe an approach where polyplots provide a more compact status of each variable. The length and direction of vectors attached to vertices of p-dimensional polygons indicate the state of each variable. To avoid Type I error distortion, the Hotelling T^2 statistic is placed near the polyplots. While these plots are more compact than separate charts, it should be noted that reliance on individual charts may not help with diagnosis in the instance of special relationships between variables as demonstrated earlier in Figures 1-1 and 1-2.

Pattern Recognition

Run rules developed by the Western Electric Company in 1956 provide decision rules for detecting non-random patterns on univariate Shewhart control charts that may suggest a problem even though the plots are within control limits (Montgomery, 1991). Chih and Rollier (1994) adapt such an approach to the bivariate case. Using simulation, specific types of shifts are induced in various combinations between the two variables.

The range of values for the T^2 statistic are divided into zones. Each combination of shift is characterized by the distribution of T^2 observations falling into each zone. Exploration of the generalizability of the method to p-variables would be of interest -- a disadvantage could be that extensive characterization of expected types of shifts over many variables would be required.

Neural Nets

When statistical properties of a problem are unknown, decision functions may be determined by training (Gonzalez and Woods, 1992). Neural networks attempt to mimic the structure of the brain by forming many layers of interconnected nodes that receive inputs and apply transformations or weights to the inputs and propogate the transformed value as node outputs. Transformation parameters are determined by presenting a known input and desired result at the input and output sections of the network, respectively. "Training" algorithms requiring many iterations are used to determine parameters that result in a suitably small error rate in providing the correct output for the known input. While much of the original work focused on creating learning machines, as well as speech, handwriting, and other pattern recognition abilities, Stern (1996) (with discussion) summarized their application in applied statistics, noting successes in time series prediction, classification, and regression applications. One application of a neural net provided results comparable to principal components. Stern (1996) noted that neural nets seem best suited to predictive problems when large datasets and substantial training

is available, but noted the disadvantage that neural nets do not provide insight into relationships between variables.

In the discussion of Stern's (1996) work, De Veaux and Ungar emphasized the need for determining what the goal is in deciding whether or not to use neural nets; does one only want to make observations/predictions, or is it better to try to achieve understanding of the process phenomenon through an empirical model? Because of highly interconnected structures with many levels, neural nets are not helpful for the latter goal. Stated advantages are: resilience to data collection errors; resistance to outliers; no requirements for equation forms; less sensitive to multicollinearity; and the flexibility to incorporate models in portions of the network. Disadvantages include: no prediction intervals indicating uncertainty; less process understanding; and requirements for large amounts of historical data and net training. The impact of data and training requirements could be felt time after time when one considers that manufacturing processes are often "tweaked," to incorporate improvements. Continual retraining of the net may be required.

As in the other multivariate methods mentioned, determination of cause may also prove to be a challenge since the inputs are passed through the net and resolved into a single statistic for comparison against a threshold.

Runger (1996a) noted that since neural nets are not based on theory that requires independent observations (with respect to time), they may be robust to the presence of autocorrelation.

Empirical Methods

Willemain and Runger (1996) noted several reasons behind a trend away from relying on charts that assume underlying normality in the data: increased automation is providing more data which may be used as empirical reference distributions; as control charts enjoy wider use, more non-normal data is encountered; statistics other than the mean with non-normal or unknown distributions are being charted; when process run times are long, individuals charts are used so that the Central Limit Theorem does not apply; and more data per unit time requires widening control limits to maintain tolerable false alarm rates per unit time.

Willemain and Runger (1996) explored the use of historical data and order statistics as empirical reference distributions, concluding that when thousands of observations are available, performance is very similar to charts that use underlying distribution theory. For symmetrical two-sided charts, they recommend that the number of observations required to form the empirical reference distribution be no lower than four times the desired in-control ARL.

Montgomery *et al.* (1993) used empirically derived principal component scores to define control regions in order to make the procedure less sensitive to autocorrelation.

The "training" of a neural net mentioned previously may also be considered an empirical method. In general, it appears the empirical approach is promising when autocorrelation is present, or when variable distributions are unknown, or are known to

violate assumptions required for other techniques. A disadvantage is that large numbers of historical observations are required.

Cross-Correlation vs. Auto-correlation in Statistical Process Monitoring

Correlation refers to relationships between variables within an observation vector while autocorrelation refers to relationships between observations of the same variable through different time periods. In a n x p matrix where the n rows represent the observations (trials) and the p columns represent the variables, correlation could be observed between variables in the same row, and autocorrelation could be observed between rows in the same column.

For multivariate quality control, correlation between variables within an observation is desirable. Hawkins (1991) states "although one could monitor the process using separate charts of the variates, to the extent that these measurements are mutually correlated, one will obtain better sensitivity using multivariate methods that exploit the correlations." Figure 2-1 shows a hypothetical control ellipse, "B", for two independent normally distributed variables. The rectangular region "A" represents a joint control region formed by the use of individual control charts on each variable. The area outside of rectangle "A", but inside ellipse "B" represents values for which one of the simultaneous individual charts would indicate a shift, when in fact, the correct joint control defined by ellipse "B" does not indicate a problem. This demonstrates the distortion of Type I error that occurs when using several individual control charts in a

multivariate situation. Even when the variables are independent, there is some benefit to using multivariate techniques to avoid this distortion.

The value of multivariate techniques really shows when relationships exist between the variables of interest. Figure 2-2 shows the control ellipse for two positively correlated variables

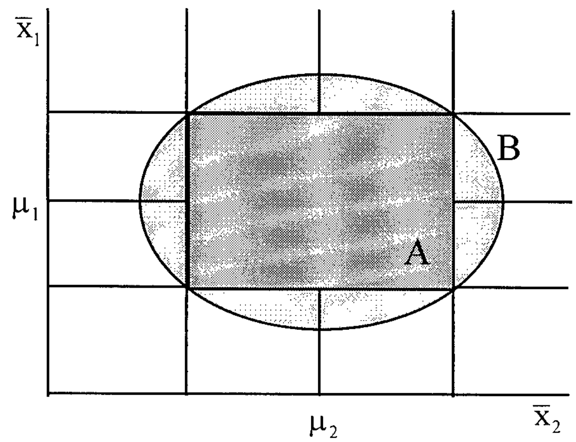


Figure 2-1. Control ellipse for two independent variables
(Adapted from Montgomery, 1991)

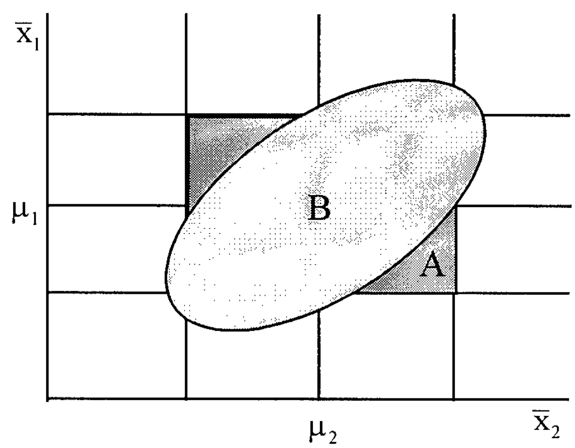


Figure 2-2. Control Ellipse for two positively correlated variables.
(Adapted from Montgomery, 1991)

Points inside the rectangle "A" but outside the ellipse "B" violate the usual relationship between the variables, but would plot inside the control limits on the individual control charts. Figure 2-2 demonstrates a real strength of multivariate control methods -- when relationships exists between the variables of interest, detection of unusual events is improved. While correlation between variables is desirable, the same cannot be said for autocorrelation between observations on the same variable.

Autocorrelation in the data may be present from any one of several causes. Montgomery and Mastrangelo (1991) mention:

> In discrete parts manufacturing, the development of sensing and measurement technology has made it possible, in many cases, to measure critical dimensions on every unit produced. Sensors are also widely used in the chemical and process industries for tanks, reactors, and material streams. All manufacturing processes are driven by inertial elements, and when the frequency of sampling becomes short relative to the process time constant the sequence of process observations will be autocorrelated.

Montgomery and Peck (1992) state "a primary cause of autocorrelation in regression problems involving time series data is failure to include one or more important regressors in the model." It's becoming not only possible, but likely, that some degree of autocorrelation will be present in industrial process data.

In univariate quality control, autocorrelation has undesirable effects. Harris and Ross (1991) and Montgomery and Mastrangelo (1991) show that positive autocorrelation results in more false alarms than would be expected if standard control charts based on the assumption of uncorrelated data are applied. Harris and Ross point out the practical implication of this problem -- the increased false alarms result in either abandoning the

use of control charts, or widening control limits to reduce the false alarms (which also decreases the sensitivity of the control chart to real process shifts). Harris and Ross also explore the alternative of fitting an ARIMA model to the data, then using a control chart on the residuals. In this scenario, positive autocorrelation makes it difficult to detect unusual events. The same analysis shows improved detection when the observations are negatively correlated. Noting that "natural" negative autocorrelation seemed to improve detection, and that industrial data is seldom negatively correlated, Keats and Shlaes (1994) propose a transformation to induce negative autocorrelation into positively autocorrelated data -- while event detection is improved, the procedure also exhibits increased false alarms. Montgomery and Mastrangelo (1991) propose a moving centerline Exponentially Weighted Moving Average (EWMA) that works "reasonably well" when the observations are positively autocorrelated at low lags and if the process drifts moderately slowly.

It is more difficult to categorize the effect of autocorrelation in multivariate quality control. Mastrangelo, Runger and Montgomery (1996) provide an excellent discussion of this issue for two types of multivariate procedures. They show the Hotelling T^2 "attenuates the autocorrelation in the process data." Their proof shows the lag one autocorrelation of the T^2 statistic is

$$\rho = \sum_{i=1}^{p} \rho_i^2 / p \qquad \text{(2-\_)}$$

where ρ_i is the lag one correlation of the ith component of $\mathbf{x}$. Because $-1 < \rho_i < 1$, ρ will always be smaller than the individual ρ_i. The presumption is that the T^2 is less sensitive to autocorrelation than univariate control techniques. For Principal Components Analysis (PCA) they discuss results of Joliffe (1986) and Jackson (1991) that show traditional assumptions of normality and independence are not essential in order for PCA to be effective in a descriptive or exploratory manner (although these assumptions are required if statistical inferences are to be made). Mastrangelo *et. al.* conclude that PCA with correlated observation vectors [autocorrelation] indicates process upsets remarkably well.

The effect of autocorrelation on the performance of multivariate control applied to regression adjusted variables (Hawkins, 1991) has not been assessed. Hawkins assumed that "the successive measurement vectors obtained over time are mutually independent." The reason for this assumption has to do with the properties of the least squares estimates for the regression parameters. One of the fundamental assumptions in linear regression is that error terms are uncorrelated (Montgomery and Peck, 1992). When this assumption is violated, the least squares regression coefficients are no longer minimum variance estimates, and MS_E may seriously underestimate σ^2, giving false impressions of accuracy and possibly indicating one or more regressors may be important when they are not (Montgomery and Peck, 1992). Presumably, the presence of autocorrelation could drive a poor model fit and adversely impact regression based control methods proposed by Hawkins. This may or may not happen since Montgomery and Peck state that it is sometimes possible to eliminate apparent autocorrelation by selecting the correct

regressors for the model. When autocorrelation is present due to true inertia in the process, then it is less likely that proper variable selection will fix the problem. Even if the problem remains, it's effect on shift detection is unclear -- the regression adjusted variables may still be analyzed using a T^2 or other procedure that may not be as sensitive to autocorrelation in the multivariate case.

There's a history of problems driven by autocorrelation in the univariate case, and effects of autocorrelation in multivariate control are largely unexplored. As a minimum, one needs to be alert to the presence of autocorrelation in the variables being monitored.

Summary

This literature review provides a sense of both the age of founding ideas yet the newness of application and resurgence of research in multivariate statistical process monitoring. Techniques under consideration vary widely, from direct extensions of univariate statistical process monitoring, to new application of other multivariate analysis methods to the process monitoring application, and to the generation of empirical methods when reference distributions are unknown.

The following investigation focuses on the researcher's interest in tailoring techniques to what is known about the process structure and anticipated assignable cause, assuming that some knowledge about the process has been gained. Regression adjustment methods show the potential for flexibility in handling non-linear relationships between data, and differing process models, though they rely heavily on assumptions that process upsets change the relationships between variables.

CHAPTER 3

MULTIVARIATE STATISTICAL PROCESS MONITORING AND DIAGNOSIS USING GROUPED REGRESSION ADJUSTED VARIABLES.

Introduction

A common theme among several of the existing multivariate statistical process monitoring (MSPM) methods is the recommendation that process knowledge should be used to select a suitable monitoring procedure. The first consideration involves considering any unique structure in the process itself. For example, Mortell and Runger (1995) propose a model and control strategies for multiple stream processes to detect shifts in the mean of all streams vs. a shift in a single stream of a multi-head filling machine. Hawkins (1993) modified his original regression adjustment approach to be more suitable for monitoring cascade (sequential value added) processes.

Other considerations involve expectations concerning the nature of a potential process upset: 1) whether a shift is likely to occur in one, several, or all of the quality characteristics being monitored; 2) whether the relationships between quality characteristics are maintained under shift conditions (model-fixed) or changed (model-void) (Runger, 1996b); and 3) whether a shift is likely to occur in any direction within the region of interest or towards a known out-of-control state (Healy, 1987). The possible combinations of these factors mounts quickly, and no single monitoring method emerges as an overall superior approach.

This chapter identifies a potentially common multivariate process monitoring scenario and extends the idea of using process knowledge to determine an appropriate control statistic for assignable cause detection and identification. Some of the material in

Chapter 2 is summarized here to provide motivation for the proposed technique.

Background

Hotelling (1947) formulated a multivariate extension to the univariate Shewhart chart. Assuming that p quality characteristics are jointly distributed as p-variate normal and that random samples of size n are collected, Hotelling's chart signals a statistically significant shift in the mean when

$$\chi_i^2 = (\mathbf{X}_i - \mu_0)' \Sigma^{-1} (\mathbf{X}_i - \mu_0) > h_1 \quad (3\text{-}1)$$

where h_1 is the selected control limit. Most often, Σ and μ_0 are unknown and must be estimated by $\overline{\mathbf{X}}$ and $\mathbf{S}$ using historical data {when these are substituted into (3-1), the resulting expression is often referred to as Hotelling's T^2}. When μ and Σ are known [or estimated from a relatively large number of preliminary samples] the upper control limit for the χ^2 chart [T^2] is often set to UCL = $\chi^2_{\alpha,p}$. The average run length (ARL) properties of the χ^2 chart depend only on the distance of the shift from the target mean vector, μ_0. This distance is given by

$$\lambda_\mu = \sqrt{(\mu - \mu_0)'\Sigma^{-1}(\mu - \mu_0)} \quad (3\text{-}2)$$

From this point forward, we assume the common practice that variables are standardized to unit normal with the zero vector as the target mean. Under this assumption, the covariance matrix becomes the correlation matrix, and the distance expression in equation (3-2) reduces to

$$\lambda_\mu = \sqrt{\mu'\Sigma^{-1}\mu} \tag{3-3}$$

Under shift conditions, the χ^2 statistic follows the non-central chi-square distribution with the non-centrality parameter equal to λ_μ^2. Lowry and Montgomery (1995) and Wierda (1994) give additional guidance for adjustments to the sample statistic and reference distribution depending on the control phase (Phase I - retrospective test for control initiation versus Phase II - continued monitoring) and the sample size (a single observation on each characteristic, or the sample mean of several observations on each characteristic).

Multivariate extensions of cumulative sum (CUSUM) and Exponentially Weighted Moving Average (EWMA) are also available. Crosier (1988) proposed the "MCUSUM" which places all the univariate CUSUMs into a vector (which is shrunk towards zero in the absence of significant deviations from the multivariate mean) then computes a length on this vector. Pignatiello and Runger (1990) introduced an alternate formulation of the multivariate CUSUM called "MC1" that accumulates distances of the sample averages from the target mean and squares this accumulated quantity. Lowry *et.al.* (1992) described a multivariate EWMA that performs as well as the best MCUSUMs, but is considered easier to formulate and possesses predictive properties.

The procedures mentioned thus far have the important property of directional invariance; that is, the average run length (ARL) to detect a shift depends only on the distance [equation (3-3)] of the shift from the multivariate mean (this is true for the MEWMA only as long as the weights used in each univariate EWMA are equal). This

directional invariance property is of special importance when characterizing a new process, or when available process knowledge suggests that shifts may occur in virtually any direction away from the target mean. In other cases, it is possible and may be desirable to increase a control scheme's sensitivity by using knowledge of possible upset mechanisms to "aim" or sensitize the control procedure. Intuitively speaking, why expend resources monitoring regions towards which the process mean is unlikely to shift? Several techniques are available that use process knowledge to sensitize shift detection and/or improve diagnosing assignable cause.

Mandel (1969) introduced the idea of a regression control chart to adjust the quality characteristic of interest (post office man-hours expended) to account for an outside covariate (mail volume) -- parallel control limits were based on a constant times residual mean squared error (versus using the regression confidence limits). In this scenario, the man-hours mean is variable, and it is only important to identify offices or changes at a specific office that violate the "usual" relationship between man-hours and mail volume -- "model fixed" changes in mean man-hours were expected and not considered important by management. Note that the property of directional invariance for the control mechanism is deliberately eliminated.

Healy (1987) proposed a CUSUM based on a linear combination of the original X_i (reducing to a univariate CUSUM procedure) that is designed to discriminate between the in-control multivariate mean and a specific value of the multivariate mean that

corresponds to an out-of-control or undesirable value. If shifts in other directions are likely, Healy recommends adding CUSUMs of linear combinations that are orthogonal.

Hawkins (1991) synthesized Mandel's regression adjustment, with Healy's approach for maximizing sensitivity to a structured shift, using the additional assumption that a shift would occur in only a single variable. The procedure involves a control statistic for each variable that is based on the residual obtained from comparing each observation to the predicted value from a regression of the jth variable on all other j-1 variables. Hawkins (1993) modified his technique for application in cascade processes by using only "upstream" observations as the independent variables in the regression adjustment. Hawkins showed these procedures are designed to detect model void types of shifts and, as such, have directionally variant ARL performance. In contrast to Hawkins' use of simultaneous univariate CUSUMs, Timm (1996) applies "stepdown" finite intersection tests (FIT) to the same regression adjusted variables.

Zhang's (1984, 1985) cause selecting control charts incorporate a two-step approach that monitors incoming quality and also monitors outgoing quality adjusted for the incoming quality. The adjustment may be made by theoretical or empirical models. The combinations of possible signals in each chart help to diagnose whether incoming, outgoing, or both measures contain assignable cause. Wade & Woodall (1993) showed that using regression prediction limits instead of confidence limits provided more predictable ARL performance in this procedure.

Runger (1996b) noted applications in which shifts in the mean may only occur in a subset of variables. For example, the manufacturing of magnetic tape consists of wet processes that coat the web with a magnetic ink and dry processes that slit and test the tape. Typically, an assignable cause only affects the mean of either the wet or dry variables, although the variables are not necessarily independent. Runger proposed the U^2 chart, a projection-based method designed to be most sensitive to shifts in a specific (guided by process knowledge) subspace of the full variable set.

Principal components (Hotelling, 1933) decomposes data into orthogonal linear combinations based on the eigenvectors of the ordered (greatest to least) eigenvalues that describe the present variability in the data. The number of principal components available is equal to the number of variables. Monitoring methods based on a full set of principal components retain the property of directional invariance. Jackson (1980) notes that most process variability is often captured in the first few principal components, and that the task of monitoring a large number of variables can be reduced into monitoring a few principal components. Scranton *et. al.* (1996) showed the improvement in ARL performance that can be obtained by using an MEWMA on a subset of principal components. The loss of directional invariance depends on the composition of the retained principal components. In the example used by Scranton *et. al*. (1996), the first three components were used which did not contain any function of two out of the original ten variables. The demonstrated control procedure would be insensitive to shifts in those

two variables. The likelihood and impact of missing a structured shift is an important factor in selecting any of the procedures described.

Grouped Regression Adjustment

Extending the idea of tailoring a control mechanism to the process structure and expected structured shifts, we may hypothesize a likely scenario not yet addressed by the aforementioned techniques:

1) For a process that may be characterized as adding value sequentially at each process step, there may be several (and varying numbers of) quality measures of interest at each step [in contrast to the $Y_1, Y_2, ..., Y_j$ single measure per step ordering of Hawkins (1993) and the one variable at a time Stepdown Finite Intersection Test of Timm(1996)]. In this type of process, a shift in an early subset of variables will propagate into downstream subset measures even though nothing may be wrong at those later stages [in contrast to Runger (1996b) in which only a distinct subset of variables is affected by an assignable cause]. For example, an assignable cause at the first step resulting in a mean shift from $\mathbf{0}$ to μ_1 will also induce a model-fixed mean shift of $\beta'\mu_1$ (where β contains the coefficients from regressing $\mathbf{X}_2$ on $\mathbf{X}_1$) into the variables of the second step (and so on), even though nothing may be wrong with the process at later steps. Hawkins refers to this downstream shift propagation as "the cascade property."

2) The variable subsets associated with each process step can form sequences (e.g. input measures, then several groups of process step measures, then output measures) -- an extension to the input/output cause selecting charts proposed by Zhang (1984, 1985). Zhang's treatment of larger multivariate scenarios primarily allows several input variables to be considered in the adjustment of a single output variable. More than one output variable would be handled in separate procedures. In the proposed scenario, it is desirable to simultaneously monitor several process measures simultaneously, after adjusting for incoming quality from the previous step.

3) If process knowledge suggests model "void" types of shifts, then relying on a small subset of principal components may be risky as the first few components are designed to be most sensitive to model "fixed" variability in the system. In cases where the principal components are composed of variables across many process steps in the cascade, determining assignable cause may be difficult.

The following notation is used to define the proposed procedure. For p total variables partitioned into k groups, let p_i for i = 1...k denote the number of variables within a group. Let the p_1 x 1 vector $\mathbf{X}_1$ denote the variables in the first group. Furthermore, let the p_i x 1 vector $\mathbf{X}_{i \cdot 1,2,\ldots,i-1}$ denote the variables in group i adjusted for the variables in the preceding groups. For example, $\mathbf{X}_{2 \cdot 1} = \mathbf{X}_2 - \Sigma_{21}\Sigma_{11}^{-1}\mathbf{X}_1$. Let $\mathbf{U}_i$ represent the control statistic [on which a T^2, MCUSUM, or MEWMA is applied independently to the elements of the i^{th} group]. The proposed control statistics are:

$$\mathbf{U}_1 = \mathbf{X}_1 \qquad \mathbf{U}_2 = \mathbf{X}_{2 \cdot 1} \qquad \mathbf{U}_3 = \mathbf{X}_{3 \cdot 1,2} \qquad \cdots \qquad \mathbf{U}_k = \mathbf{X}_{k \cdot 1,2,\ldots,k-1}$$

Note that when all $p_i = 1$, this method reduces to the approach of Hawkins (1993).

Intuitively we see that each successive group has the effects of all variables in previous groups removed. An important advantage is that these statistics are independent of each other. A proof of this result is given in Appendix 3.A. Independence of these statistics allows us to design the Type I error for the joint procedure by setting the α for each group such that the joint procedure Type I error is $\alpha = 1 - (1 - \alpha')^k$ and the desired joint in-control ARL is $1/\alpha$ (using T^2 charts).

For the first group, the mean vector (the zero vector under H_0) and correlation matrix are formed in standard fashion using the variables in $\mathbf{X}_1$. The mean vector and covariance matrices for the subsequent groups are conditional on the earlier groups (see Mardia *et. al.*, 1979, p.63). For example,

$$\mu_{2.1} = \mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\,\mu_1 \tag{3-4}$$

$$\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \tag{3-5}$$

where $\mathbf{X}$ is partitioned into $\mathbf{X}_1$ containing all variables in preceding groups and $\mathbf{X}_2$ containing the variables in the group under consideration (downstream variables excluded). Also, Σ is partitioned accordingly.

Due to the conditional nature of the proposed control statistics (with the exception of that applied to the first group which maintains directional invariance), ARL performance is highly dependent on the partitioned covariance matrix as well as the shift size and direction. An example based on actual process data serves to illustrate the advantages of the proposed approach compared to the blanket use of a single χ^2 chart on $\mathbf{X}$. While the χ^2 chart comparison is used for simplicity, MCUSUM or MEWMA techniques could also be applied to the proposed $\mathbf{U}$ statistics for better overall ARL performance.

Example

This example is a modification of the cotton spinning example used by Duncan (1986) and Hawkins (1993) {see Appendix 3.B for explanation of modifications}. The following quality characteristics are of interest:

X_1 = Fiber Fineness
X_2 = Fiber Length
X_3 = Fiber Strength
X_4 = Skein Strength

For demonstration's sake, Hawkins assumed the process model that these four variables satisfy the cascade property, with shifts in any variable propagating to downstream

variables. From the variable descriptions in Duncan (1986, p. 835 -- the source of Hawkins' example data) it is apparent that Fiber Fineness, Fiber Length, and Fiber Strength are innate properties of a fiber that is loaded into the loom for spinning. Skein Strength is the strength of the weave that is produced as an output from the loom. An assignable cause in the production of the input fiber would be likely to affect several if not all of the "input" measures. These, in turn, could influence X_4, satisfying the cascade property between steps. Of course, an assignable cause associated with the loom could shift the output measure. These assumptions support an input/output grouping of the variables. A second output measure X_5, "Skein Stretch", is fabricated to support the proposed process scenario [see Appendix 3.B for details].

For an in-control ARL of 200, the joint procedure Type I error = .005. To maintain this, α is set at 0.0025 for each of the two groups. For $\mathbf{U}_1$ the χ^2 chart control limit is $\chi^2_{.0025,3} = 14.32$. For $\mathbf{U}_2$ the χ^2 chart control limit is $\chi^2_{.0025,2} = 11.98$.

Case 1 - Shift in the spinning process. Suppose that the input fiber characteristics remain in a state of control, but a problem with the spinning process occurs such that the mean of X_4 (X_5) is decreased (increased) by one standard deviation such that $\mu'_X = [0,0,0,-1,1]$. The means of the proposed control statistics then are $\mu'_{U1} = [0,0,0]$ and $\mu'_{U2} = [-1, 1]$. This is a model-void shift that regression adjustment procedures detect well. Using the correlation matrix in Appendix 3.B with equations

(3-5) then (3-3), the distance of the shift, λ, is 2.083 (see Appendix 3.C for verification of shift distance and non-centrality parameter calculations). From the non-central chi-square distribution, the probability of obtaining a point above the $\mathbf{U}_2$ control limit is 0.11713 (See Appendix 3.D for SAS code/output listings used to obtain values under non-central chi-squared distribution). The ARL for the joint procedure under this shift condition is 1 / [1-(.9975)(.88287)] = 8.380.

If the entire variable set is monitored with the traditional chi-square chart, the original correlation matrix is used directly in (3-3) to determine the shift distance, λ, of 2.083. The control limit is $\chi^2_{.005,5}$ = 16.75. Under this shift condition, the probability of observing a point above the control limit is 0.09178. The ARL for this procedure is 1/0.09178 = 10.896.

While the assumed shift is model-void between the groups, the general relationship (negative correlation) between X_4 and X_5 is maintained. Consider another model-void shift where this relationship is also violated such that μ'_X = [0,0,0,-1,-1]. For this shift, the ARL of the proposed procedure is 2.73, while the ARL of the traditional chi-square chart is 3.47.

In Case 1, the shift detection is faster using the proposed procedure since the variance of residuals contained in $\mathbf{U}_2$ is much smaller than the variance of the original [X_4, X_5], and this provides a stronger "signal to noise" ratio. It is likely that the signal would correctly be in $\mathbf{U}_2$ indicating a problem with the spinning process {though a false alarm in $\mathbf{U}_1$ is still remotely possible}.

A Principal Components Analysis using SAS PRINCOMP produced the components in Table 3-1 and the eigenvalue analysis in Table 3-2 (See Appendix 3.E for SAS Code/Output Listing). Although the first two (three) principal components capture 82% (91%) of the variability in the dataset, the loadings in the principal components highlight a potential problem with their use. Since at least one variable in both the input (X_1 - X_3) and output (X_4, X_5) variable sets has a substantial loading, a signal in a principal component would be difficult to interpret -- is the assignable cause due to the input fiber properties, or to a problem in the spinning process, or both? This problem is especially severe in the first principal components where all variable loadings are approximately equal in magnitude. If process knowledge of likely shift structures suggested shifts in the direction of principal components, then a combined approach of monitoring with PCA, and diagnosis with $\mathbf{U}_k$ may be advisable.

TABLE 3-1. Principal Components of Cotton Fiber Data

Variable	1st PC	2nd PC	3rd PC	4th PC	5th PC
X_1	-.402885	0.539649	0.168018	0.613428	0.376746
X_2	0.402673	0.549622	0.491714	-.020977	-.541800
X_3	0.357925	-.575535	0.415586	0.605516	0.035894
X_4	0.554037	0.185275	0.144552	-.294608	0.742312
X_5	-.490379	-.202799	0.732380	-.412095	0.110450

TABLE 3-2. Eigenvalue Analysis of Cotton Fiber Data Correlation Matrix.

	Eigenvalue	Proportion	Cum Proportion
Prin1	2.74064	0.548127	0.54813
Prin2	1.37028	0.274055	0.82218
Prin3	0.42228	0.084455	0.90664
Prin4	0.33247	0.066495	0.97313
Prin5	0.13434	0.026868	1.00000

Case 2 - Shift in the input fiber characteristics. Suppose the spinning process was functioning correctly, but that the input fiber is off target. Possible causes of this situation may be selection of incorrect material during loom loading, or loading of material thought to be correct, but contains off-target properties due to problems in its manufacture. Assume a shift such that $\mu_1' = [1,-1,-1]$. Furthermore, assume that μ_2 undergoes a model-fixed shift due to the cascade property in the process. This shift is determined by the following regression equations obtained using the original correlation matrix:

$$\begin{aligned} X_4 &= -0.426X_1 + 0.712X_2 + 0.054X_3 \\ X_5 &= \ \ 0.319X_1 - 0.534X_2 - 0.040X_3 \end{aligned} \qquad (3\text{-}6)$$

Applying equations (3-6), $\mu_2' = [-1.192, 0.893]$. The complete shift structure in **X** is $\mu'_X = [1,-1,-1,-1.192, 0.893]$. From equation (3-4), the mean of $\mathbf{U}_2$ remains $[0,0]'$. The vector μ_1 and matrix Σ_{11} are used in (3-3) to calculate the $\mathbf{U}_1$ shift distance, λ, of 1.4536. From the non-central chi-square distribution, the probability of obtaining a point above the $\mathbf{U}_1$ control limit is 0.02804. The ARL for the joint procedure under this shift condition is 1 / [1-(.97196)(.9975)] = 32.819.

Using a traditional chi-square chart on all **X**, the original correlation matrix is used with (3) to determine the shift distance, λ, of 1.4536. The control limit is $\chi^2_{.005,5} = 16.75$. Under this shift condition, the probability of observing a point above the control limit = 0.03235. The ARL for this procedure is 1/0.03235 = 30.912.

In this case, the cascade property induces changes in downstream variables as well, and all contribute to the χ^2 chart on **X** helping it to signal slightly faster than the joint procedure involving $\mathbf{U}_k$. While the faster signal is desirable, diagnosis of the assignable cause using decomposition methods on the full set χ^2 statistic may be misleading. For example, Murphy's (1987) discriminant-analysis-based approach partitions **X** into $\mathbf{X}_1$, containing suspect variables, and $\mathbf{X}_2$, containing all other variables. Defining T_p^2 as the T^2 statistic on the full set of variables, and $T_{p_1}^2$ as the T^2 statistic on the suspect set of variables, Murphy defines the difference statistic $D = T_p^2 - T_{p_1}^2$. When D is large, the hypothesis that the p_1 subset caused the signal is rejected. Placing X_4 and X_5 into the suspect set $\mathbf{X}_1$, D = 2.113 - 1.428 = .685. We note that $T_{p_1}^2$ is almost as large as T_p^2, indicating that X_4 and X_5 are contributors, though we know nothing is wrong with the loom in this scenario. This problem is not due to Murphy's procedure, but rather that the procedure is applied directly to **X** without accounting for shift propagation in the process. In contrast, signals in the $\mathbf{U}_k$ directly indicate a problem within the associated variable subset.

Case 3 - Composite Shift. Table 3-3 provides the results from combining the shifts in cases 1 and 2. Again, the cascade property in the process allows the input material change to induces further changes in X_4 and X_5 above and beyond the process shift, contributing to a slightly faster signal in the χ^2 chart on **X**. Decomposition is likely

TABLE 3-3. Comparison of χ^2 chart on $\mathbf{U}_k$ vs. on $\mathbf{X}$ for composite shift case.

Stat	Mean	α	UCL	λ	λ^2	$\chi^2_{\alpha,p,\lambda}$	ARL	Joint ARL
$\mathbf{U}_1$	[1,-1,-1]	.0025	14.32	1.4536	2.113	0.97196	35.66	7.05
$\mathbf{U}_2$	[-1,1]	.0025	11.98	2.0833	4.340	0.88287	8.54	
$\mathbf{X}$	[1,-1,-1, -2.192, 1.893]	.0050	16.75	2.5399	6.451	0.82420	5.69	

to correctly show causes in both groups; however, the "trouble" in the second group is likely to be overstated.

Other Impacts Associated with Amount of Grouping

The partitioning of p variables into k groups in this procedure is not arbitrary. It is driven by the structure of the process under consideration and associated potential upset mechanisms. It would be undesirable for the amount of grouping alone to significantly influence the ARL performance. Table 3-4 investigates various levels of groupings of p = 12 variables (see Appendix 3.F for SAS code/output listings). The key assumption in Table 3-4 is that a shift occurs in the mean of a single group of variables. As the number of groups increases, control limits must be raised to control Type I error. A higher number of groups results in a smaller number of variables in each group. The fewer degrees of freedom within a group reduces the control limit for a group. These effects act in opposition, tending to balance each other for small shifts. There are minor differences in ARL performance for moderate to large shifts.

TABLE 3-4. Average Run Lengths of χ^2 Chart Combined Procedure Involving k charts (groups) with p_i variables each ($p = 12$).

$k\{p_i\}$	6{2}	4{3}	3{4}	2{6}
UCL	14.175975	15.78961	17.32726	20.24635
λ	ARL			
0	200.000	200.000	200.000	200.000
0.5	170.698	169.996	169.316	168.053
1	99.104	100.057	100.510	100.645
1.5	41.723	43.771	45.287	47.319
2	16.459	17.790	18.892	20.614
2.5	7.185	7.846	8.429	9.416
3	3.670	3.988	4.279	4.795

Summary and Conclusions

Extending the idea that process knowledge should be used to select a suitable monitoring method, this paper introduces control statistics appropriate for the monitoring and shift diagnosis of a cascade process involving several (and varying numbers of) quality characteristics at each process step. While a shift detection in the first group of variables depends only on the statistical distance of the shifted mean from the target mean, detection in subsequent groups is very much a function of the correlation structure between variables as well as the structure of anticipated shifts. In cases where the shift is in any but the first group, shift detection may be faster using the proposed procedure since the variance of residuals contained in $\mathbf{U}_2 \ldots \mathbf{U}_k$ is smaller than the variance of the original $\mathbf{X}_2 \ldots \mathbf{X}_k$, providing a stronger "signal to noise" ratio.

In some cases, a signal in a subset is enough to identify assignable cause. In those cases where it is not, then this subset forms a reduced starting set for adaptations of

diagnostic methods proposed by Murphy (1987), Chua and Montgomery (1992), and Mason, Tracy, and Young (1995). Since the performance of these methods depends, in part, on the number of variables involved, the benefit of this approach is starting the identification process with a smaller and "corrected" subset of variables. It's important to note that using these diagnostics on statistics that rely on the full set of **X** (such as decomposing a single T^2 that contains all variables in **X** metric) could lead to erroneous diagnosis since shifts are propagated downstream in the process structure under consideration.

Should the study of a process indicate that another monitoring technique provides better signaling performance for anticipated shift structures, the proposed statistics retain diagnostic value under the assumed process model.

APPENDIX 3.A

Proof of Independence Among Proposed Control Statistics

At the current process step, consider a partition of $\mathbf{X}$ into $\mathbf{X}_1$ (an n x p_1 matrix containing n observations of variables in all preceding steps), $\mathbf{X}_2$ (an n x p_2 matrix containing variables in the current step) and $\mathbf{X}_3$ (downstream variables), the $\mathbf{U}_i$ partition $\mathbf{X}$ such that:

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 \\ \mathbf{U}_2 \end{pmatrix} = \begin{bmatrix} \mathbf{I}_{p_1} & \mathbf{0} \\ -\mathbf{B} & \mathbf{I}_{p_2} \end{bmatrix} \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}$$

where $\mathbf{B}$ is a p_2 x p_1 matrix, the rows of which contain parameter estimates from a regression of each variable in $\mathbf{X}_2$ on all variables in $\mathbf{X}_1$. From familiar multivariate normal theory (see Tong, p.32), If $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$ and $\mathbf{U} = \mathbf{CX} + \mathbf{b}$, then $\mathbf{U} \sim \mathcal{N}_p(\mathbf{C}\boldsymbol{\mu} + \mathbf{b}, \mathbf{C\Sigma C}')$. Letting

$$\mathbf{C} = \begin{bmatrix} \mathbf{I}_{p_1} & \mathbf{0} \\ -\mathbf{B} & \mathbf{I}_{p_2} \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

then

$$\Sigma_{\mathrm{U}} = \begin{bmatrix} \mathbf{I}_{p_1} & \mathbf{0} \\ -\mathbf{B} & \mathbf{I}_{p_2} \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{p_1} & -\mathbf{B} \\ \mathbf{0} & \mathbf{I}_{p_2} \end{bmatrix}$$

which reduces to:

$$\Sigma_{\mathrm{U}} = \begin{bmatrix} \Sigma_{11} & (\Sigma_{21} - \mathbf{B}\Sigma_{11})' \\ \Sigma_{21} - \mathbf{B}\Sigma_{11} & \Sigma_{22} + \mathbf{B}\Sigma_{11}\mathbf{B}' - \mathbf{B}\Sigma_{12} - \Sigma_{21}\mathbf{B}' \end{bmatrix}$$

For $\mathbf{U}_1$ and $\mathbf{U}_2$ to be independent, **B** must be chosen so that $\Sigma_{21} - \mathbf{B}\Sigma_{11} = 0$. This condition is met when $\mathbf{B} = \Sigma_{21}\Sigma_{11}^{-1}$. To complete the proof, we must show that this expression for **B** is equivalent to the least-squares parameter estimates obtained through regression. Allowing **X** to represent the regressors from $\mathbf{X}_1$, and allowing **Y** to represent the "dependent" variables in $\mathbf{X}_2$, we note that since X and Y have been standardized:

$$\Sigma_{11} = \frac{\mathbf{X'X}}{\mathbf{n-1}} \text{ and } \Sigma_{12} = \frac{\mathbf{X'Y}}{\mathbf{n-1}}$$

Substituting into the expression for **B**,

$$\frac{\mathbf{X'X}}{\mathbf{n-1}}\mathbf{B} = \frac{\mathbf{X'Y}}{\mathbf{n-1}}$$

Multiplying both sides by n-1 leaves

$$\mathbf{X'X\,B = X'Y}$$

which we recognize as the familiar least-squares normal equations for which the solution is:

$$\mathbf{B = (X'X)^{-1}\,X'Y}$$

The least-squares estimate of **B** meets the conditions for independence between $\mathbf{U}_1$ and $\mathbf{U}_2$.

APPENDIX 3.B

Modification of Hawkins (1993) Example

Hawkins (1993) illustrated the efficacy of the Y_j statistic using the cotton yarn data from Duncan (1986, p. 835). Duncan labeled the variables in the following manner:

X_1 = Skein strength
X_2 = Fiber Length
X_3 = Fiber Strength
X_4 = Fiber Fineness

Hawkins modified this so that the variable numbering more closely followed the "natural" ordering of a sequential value-added (cascade) process. In doing so, he re-labeled the variables according to

X_1 = Fiber Fineness
X_2 = Fiber Length
X_3 = Fiber Strength
X_4 = Skein Strength

We were unable to duplicate some of the basic summary statistics reported by Hawkins. For example, we found the correlation between Fiber Fineness (X_1) and Tensile Strength (X_3) to be -0.666, whereas Hawkins reported it to be -0.16. The multiple regression of X_4 on the other variables was reported to be $X_4 = -0.343X_1 + 0.606X_2 + 0.352X_3$; whereas, we found it to be $X_4 = -0.426X_1 + 0.712X_2 + 0.055X_3$. We note that the primary differences are in relationships involving X_3. This does not effect the validity of Hawkins' results (the source may simply be a difference in several of the data points entered upon which a correct analysis was made) -- it merely makes comparison difficult. To make the comparisons in this paper we use the data set found in

Duncan (1986, p. 835) but retain the labeling convention used by Hawkins that reflects the more "natural" ordering.

Furthermore, the variable descriptions in Duncan indicate that X_1 - X_3 are properties of a fiber "raw material" that is then woven into a skein (the strength of which is X_4). This makes for a natural input / output grouping of variables for use in our examples. Since a group size of one in the output grouping is overly simplistic, a second output variable "Skein Stretch" (X_5) was fabricated. This variable is posited to have a highly negative correlation with Skein Strength (X_4) as a stronger weave should stretch less and vice versa. This correlation was chosen as - 0.7; moderately strong, but not the strongest exhibited in the dataset. The relationships between X_5 and the other variables are posited to be similar in size to those of X_4 (again operating in the opposite direction), but attenuated approximately 25% to accommodate a sense that stretch may be slightly more a function of the weave process (vs. fiber characteristics) than is the strength measure.

For standardized $\mathcal{N}(0,1)$ variables, the covariance matrix is the correlation matrix. Based on the above modifications to the Duncan cotton fiber data, the correlation matrix used for this example is:

	X_1	X_2	X_3	X_4	X_5
X_1	1.000	-0.035	-0.666	-0.487	0.365
X_2	-0.035	1.000	0.041	0.729	-0.547
X_3	-0.666	0.041	1.000	0.367	-0.275
X_4	-0.487	0.729	0.367	1.000	-0.700
X_5	0.365	-0.547	-0.275	-0.700	1.000

This correlation matrix is assumed known and appropriate partitions are used in (3-4) and (3-5) to determine the correct mean vector and covariance matrix used to calculate the non-centrality parameter in (3-3) used to compute the shift condition ARL.

The raw data from Duncan (with variable numbers re-labelled to Hawkins convention) is:

Piece #	Fiber Fineness (0.1 Micrograms/Inch) X_1	Fiber Length (0.01 Inch) X_2	Fiber Tensile Strength (1,000 PSI) X_3	Skein Strength (Pounds) X_4
1	44	85	76	99
2	42	82	78	93
3	42	75	73	99
4	44	74	72	97
5	43	76	73	90
6	46	74	69	96
7	46	73	69	93
8	36	96	80	130
9	36	93	78	118
10	37	70	73	88
11	46	82	71	89
12	45	80	72	93
13	42	77	76	94
14	50	67	76	75
15	48	82	70	84
16	41	76	76	91
17	31	74	78	100
18	29	71	80	98
19	39	70	83	101
20	38	64	79	80

APPENDIX 3.C

Calculation of Non-Centrality Parameters and Shift Distances (MathCad [3.0] Objects)

$$Sig := \begin{bmatrix} 1 & -0.035 & -0.666 & -0.487 & 0.365 \\ -0.035 & 1 & 0.041 & 0.729 & -0.547 \\ -0.666 & 0.041 & 1 & 0.367 & -0.275 \\ -0.487 & 0.729 & 0.367 & 1 & -0.7 \\ 0.365 & -0.547 & -0.275 & -0.7 & 1 \end{bmatrix}$$

Correlation Matrix

$$Sig11 := \begin{bmatrix} 1 & -0.035 & -0.666 \\ -0.035 & 1 & 0.041 \\ -0.666 & 0.041 & 1 \end{bmatrix} \quad Sig12 := \begin{bmatrix} -0.487 & 0.365 \\ 0.729 & -0.547 \\ 0.367 & -0.275 \end{bmatrix}$$

$$Sig21 := \begin{bmatrix} -0.487 & 0.729 & 0.367 \\ 0.365 & -0.547 & -0.275 \end{bmatrix} \quad Sig22 := \begin{bmatrix} 1 & -0.7 \\ -0.7 & 1 \end{bmatrix}$$

Partitioned Correlation Matrix

$$Sig2\_1 := Sig22 - \left[Sig21 \cdot Sig11^{-1} \cdot Sig12 \right]$$

$$Sig2\_1 = \begin{bmatrix} 0.254 & -0.14 \\ -0.14 & 0.58 \end{bmatrix}$$

Conditional Covariance Matrix For Second Group

Shift Distances and Non-centrality Parameters:

Case 1 {Joint Procedure}:

$$muU2 := \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Shift in U2 Mean.

$$lamda2U2 := muU2^{T} \cdot Sig2\_1^{-1} \cdot muU2$$

$$lamda2U2 = 4.34$$

Non-centrality parameter (lamda-squared) for U2.

$$lamdaU2 := \sqrt{4.34}$$

$$lamdaU2 = 2.083$$

U2 Shift Distance, lamda.

Case 1 {Traditional}

$$mux := \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \\ 1 \end{bmatrix}$$

Shift in Mean of X.

$$lamda2X := mux^T \cdot Sig^{-1} \cdot mux$$

$$lamda2X = 4.34$$

Non-centrality parameter (lamda-squared).

$$lamdaX := \sqrt{4.34}$$

$$lamdaX = 2.083$$

Shift Distance, lamda.

Case 1{Joint Procedure; X4-X5 Void}:

$$muU2 := \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

Shift in U2 Mean

$$lamda2U2 := muU2^T \cdot Sig2\_1^{-1} \cdot muU2$$

$$lamda2U2 = 8.738$$

Non-centrality parameter (lamda-squared).

$$lamdaU2 := \sqrt{8.738}$$

$$lamdaU2 = 2.956$$

Shift Distance, lamda

Case 1 {Traditional; X4-X5 Void}:

$$mux := \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \\ -1 \end{bmatrix}$$

Shift in Mean of X

$$lamda2X := mux^{T} \cdot Sig^{-1} \cdot mux$$

$$lamda2X = 8.738$$

Non-centrality parameter (lamda-squared).

$$lamdaX := \sqrt{8.738}$$

$$lamdaX = 2.956$$

Shift Distance, lamda

Case 2 {Joint Procedure}:

$$muU1 := \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}$$

Shift in U1 Mean

$$lamda2U1 := muU1^{T} \cdot Sig11^{-1} \cdot muU1$$

$$lamda2U1 = 2.113$$

Non-centrality parameter (lamda-squared).

$$lamdaU1 := \sqrt{2.113}$$

$$lamdaU1 = 1.454$$

Shift Distance, lamda

Case 2 {Traditional}:

$$mux := \begin{bmatrix} 1 \\ -1 \\ -1 \\ -1.192 \\ 0.893 \end{bmatrix}$$

Shift in Mean of X

$lamda2X := mux^{T} \cdot Sig^{-1} \cdot mux$

$lamda2X = 2.113$

Non-centrality parameter (lamda-squared).

$lamdaX := \sqrt{2.113}$

$lamdaX = 1.454$

Shift Distance, lamda

Case 3 {Traditional}:

$$mux := \begin{bmatrix} 1 \\ -1 \\ -1 \\ -2.192 \\ 1.893 \end{bmatrix}$$

Shift in Mean of X

$lamda2X := mux^{T} \cdot Sig^{-1} \cdot mux$

$lamda2X = 6.451$

Non-centrality parameter (lamda-squared).

$lamdaX := \sqrt{6.451}$

$lamdaX = 2.54$

Shift Distance, lamda

APPENDIX 3.D

Computation of Probabilities Under Non-Central Chi-Square Distribution

SAS Code

```
data chi;
p1 = probchi(11.98,2,4.34);      /* Case 1, U2          */
p2 = probchi(16.75,5,4.34);      /* Case 1, Traditional */
p3 = probchi(14.32,3,2.113);     /* Case 2, U1          */
p4 = probchi(16.75,5,2.113);     /* Case 2, Traditional */
p5 = probchi(16.75,5,6.451);     /* Case 3, Traditional */
proc print data = chi;
run;
```

SAS Output Listing

```
                    The SAS System

OBS       P1         P2         P3         P4         P5

1      0.88287   0.90822   0.97196   0.96765   0.82420
```

SAS Code

```
data chi;
P0 = probchi(11.98,2,8.738);     /* Case 1, U2, X4-X5 Void    */
P1 = probchi(16.75,5,8.738);     /* Case 1, Trad., X4-X5 Void */
proc print data = chi;
run;
```

SAS Output Listing

```
                    The SAS System

OBS       P0         P1

1      0.63548   0.71158
```

APPENDIX 3.E

Principal Components Analysis (PCA) of Cotton Fiber Data

SAS Code

```
options ls = 78;
title "Principal Components Analysis of Cotton Fiber Data";
;
data cotton(type=corr);
     infile cards missover;
     _type_='corr';
     input _name_ $ x1 x2 x3 x4 x5;
     cards;
x1 1.000
x2 -0.035 1.000
x3 -0.666 0.041 1.000
x4 -0.487 0.729 0.367 1.000
x5 0.365 -0.547 -0.275 -0.700 1.000
;
proc print data=cotton;
;
proc princomp cov data=cotton;
var x1 x2 x3 x4 x5;
;
run;
```

SAS Output Listing

Principal Components Analysis of Cotton Fiber Data

OBS	_TYPE_	_NAME_	X1	X2	X3	X4	X5
1	corr	x1	1.000	.	.	.	.
2	corr	x2	-0.035	1.000	.	.	.
3	corr	x3	-0.666	0.041	1.000	.	.
4	corr	x4	-0.487	0.729	0.367	1.0	.
5	corr	x5	0.365	-0.547	-0.275	-0.7	1

Principal Components Analysis of Cotton Fiber Data

Principal Component Analysis

10000 Observations
5 Variables

Total Variance = 5

Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
PRIN1	2.74064	1.37036	0.548127	0.54813
PRIN2	1.37028	0.94800	0.274055	0.82218
PRIN3	0.42228	0.08980	0.084455	0.90664
PRIN4	0.33247	0.19813	0.066495	0.97313
PRIN5	0.13434	.	0.026868	1.00000

Eigenvectors

	PRIN1	PRIN2	PRIN3	PRIN4	PRIN5
X1	-.402885	0.539649	0.168018	0.613428	0.376746
X2	0.402673	0.549622	0.491714	-.020977	-.541800
X3	0.357925	-.575535	0.415586	0.605516	0.035894
X4	0.554037	0.185275	0.144552	-.294608	0.742312
X5	-.490379	-.202799	0.732380	-.412095	0.110450

APPENDIX 3.F

ARL Calculations for Level of Grouping Comparison (Table 3-4)

Six (6) groups of two (2) variables each:

SAS Code

```
data chi;
ARL0 = 1/(1-(probchi(14.175975,2,0))**6);
ARL05 = 1/(1-(.999164925**5)*(probchi(14.175975,2,0.25)));
ARL1 = 1/(1-(.999164925**5)*(probchi(14.175975,2,1)));
ARL15 = 1/(1-(.999164925**5)*(probchi(14.175975,2,2.25)));
ARL2 = 1/(1-(.999164925**5)*(probchi(14.175975,2,4)));
ARL25 = 1/(1-(.999164925**5)*(probchi(14.175975,2,6.25)));
ARL3 = 1/(1-(.999164925**5)*(probchi(14.175975,2,9)));
proc print data = chi;
run;
```

SAS Output Listing

```
                    The SAS System


OBS    ARL0     ARL05     ARL1     ARL15     ARL2     ARL25     ARL3
 1   200.000  170.698  99.1039  41.7232  16.4591  7.18468  3.66950
```

Four (4) groups of three (3) variables each:

SAS Code

```
data chi;
ARL0 = 1/(1-(probchi(15.78961,3,0)**4));
ARL05 = 1/(1-(.99874765**3)*(probchi(15.78961,3,.25)));
ARL1 = 1/(1-(.99874765**3)*(probchi(15.78961,3,1)));
ARL15 = 1/(1-(.99874765**3)*(probchi(15.78961,3,2.25)));
ARL2 = 1/(1-(.99874765**3)*(probchi(15.78961,3,4)));
ARL25  = 1/(1-(.99874765**3)*(probchi(15.78961,3,6.25)));
ARL3   = 1/(1-(.99874765**3)*(probchi(15.78961,3,9)));
proc print data = chi;
run;
```

SAS Output Listing

```
                    The SAS System

OBS    ARL0     ARL05     ARL1     ARL15     ARL2     ARL25     ARL3
 1   200.000  169.996  100.057  43.7712  17.7885  7.84566  3.98834
```

Three (3) groups of four (4) variables each:

SAS Code

```
data chi;
ARL0 = 1/(1-(probchi(17.32726,4,0)**3));
ARL05 = 1/(1-(.99833055**2)*(probchi(17.32726,4,.25)));
ARL1 = 1/(1-(.99833055**2)*(probchi(17.32726,4,1)));
ARL15 = 1/(1-(.99833055**2)*(probchi(17.32726,4,2.25)));
ARL2 = 1/(1-(.99833055**2)*(probchi(17.32726,4,4)));
ARL25  = 1/(1-(.99833055**2)*(probchi(17.32726,4,6.25)));
ARL3   = 1/(1-(.99833055**2)*(probchi(17.32726,4,9)));
proc print data = chi;
run;
```

SAS Output Listing

```
                    The SAS System


OBS    ARL0     ARL05     ARL1     ARL15     ARL2     ARL25     ARL3
1    200.000  169.316  100.510  45.2874  18.8921  8.42874   4.2791
```

Two (2) groups of six (6) variables each:

SAS Code

```
data chi;
ARL0 = 1/(1-(probchi(20.24635,6,0)**2));
ARL05 = 1/(1-(.99749687)*(probchi(20.24635,6,.25)));
ARL1 = 1/(1-(.99749687)*(probchi(20.24635,6,1)));
ARL15 = 1/(1-(.99749687)*(probchi(20.24635,6,2.25)));
ARL2 = 1/(1-(.99749687)*(probchi(20.24635,6,4)));
ARL25  = 1/(1-(.99749687)*(probchi(20.24635,6,6.25)));
ARL3   = 1/(1-(.99749687)*(probchi(20.24635,6,9)));
proc print data = chi;
run;
```

SAS Output Listing

```
                    The SAS System


OBS    ARL0     ARL05     ARL1     ARL15     ARL2     ARL25     ARL3
1    200.000  168.053  100.645  47.3185  20.6141  9.41595   4.79488
```

CHAPTER 4

MULTIVARIATE STATISTICAL PROCESS MONITORING INVOLVING NON-NORMAL DATA

Introduction

The form of many variables control chart statistics and their calculated Average Run Length performance depends on assumptions that the variables of interest are normally distributed. While attributes control charts deal explicitly with non-normality by using appropriate distributions (binomial for *fraction defective*, exponential for *time between events*, etc.), variables control charts used in the presence of non-normality mostly rely on the use of batch means with an appeal to the Central Limit Theorem. When process run times are long, some economy is lost in doing this since several observations must be taken from the process before a single point on the control chart is realized.

Without batching, regression adjustment techniques on individual observations in the presence of non-normality face additional problems since the non-normality may adversely affect model fit in addition to its impact on the chart used to monitor the residuals.

This chapter presents a semi-conductor manufacturing scenario that seems well-suited to monitoring with regression adjusted variables, but assumptions of normality are violated. Rather than batching observations to obtain normality, the use of Generalized Linear Models (GLM) theory (first published by Nelder and Wedderburn, 1972) to explicitly deal with the non-normality is explored.

Background

Control charts are often classified into two types: 1) *Variables Control Charts* when the quality characteristic of interest may be expressed as a number on a continuous scale; and 2) *Attributes Control Charts* when the characteristic is discrete (*e.g.* number of defects), or non-quantitative (*e.g.* conforming/non-conforming) (Montgomery, 1991). Design of *Variables Control Charts* typically rely heavily on normal distribution theory. A Shewhart chart on individual observations typically uses control limits at $\pm$ 3-sigma symmetrically about the mean. An $\overline{x}$ chart also typically uses $\pm$ 3-standard error limits symmetrically about the mean. Under the normal distribution, the probability of a point lying outside these limits is 0.0027. Hence, the in control Average Run Length (ARL) of these charts is $1/0.0027 \approx 370$ observations.

CUSUM procedures are based on sequential probability ratio tests that for many common distributions reduce to calculating cumulative sums (Healy, 1987). The ARL performance of Variables Control Chart CUSUMs is generally determined by the Markov-chain based approach proposed by Brook and Evans (1972), who recommended "discretizing" the normal distribution by dividing it into intervals and forming discrete state-transition probabilities across these intervals. ARL performance for Fast Initial Response (FIR) procedures (basically giving the CUSUM a headstart towards its control limit) and for robust procedures (such as requiring two in a row above the control limit) have been calculated by assessing the net effect of the modification on the state transition probabilities. While ARL performance for Attributes CUSUMs have been calculated by

using the appropriate distribution to determine state-transition probabilities (binomial for *fraction defective*, exponential for *time between events*, etc.), ARL performance calculations for Variables CUSUMs are primarily based on assumptions of normality. When the underlying distribution is unknown (or known to be non-normal), using batch means and appealing to the central limit theorem often satisfies the normality assumption. When process run times are long, some economy is lost in doing this, since several independent observations must be taken from the process before a single point on the control chart is realized.

ARL results for Exponentially-Weighted Moving Average (EWMA) charts published by Lucas and Sacucci (1990) also rely on Markov chain methods applied to a discretized normal distribution, though Montgomery (1991) stated that the EWMA is insensitive to the normality assumption since it may be viewed as a weighted average of all past and current observations.

In Multivariate Statistical Process Monitoring, Hotelling's T^2 statistic is derived using assumptions of p-variate normality. Healy's (1987) MCUSUM procedure reduces to a univariate CUSUM, with the implication that univariate variable CUSUM chart performance predictions (which are based on normal distribution theory) may be used. Methods compared in Pignatiello and Runger (1990) either: 1) use the result that ARL performance is determined solely by the statistical distance of the shift and use the Non-Central Chi-Square distribution to calculate ARLs; or 2) simulate ARL performance using multi-normal variates when the assumption in 1) is not met. Use of the Non-

Central Chi-Square distribution implies assumptions that the quality characteristics of interest are normally distributed.

The Chi-Square Distribution is used to model a random variable that is a quadratic form of normally distributed, standardized random variables such that (Hines and Montgomery, 1980):

$$X^2 = Z_1^2 + Z_2^2 + \ldots + Z_p^2 \tag{4-1}$$

The Non-Central Chi-Square distribution models similar quadratic forms where one or more of the means of the normally distributed random variables have a constant added such that their expected values are no longer zero (Tiku, 1985):

$$\chi'^2 = \sum_{i=1}^{p} (Z_i + a_i)^2 \tag{4-2}$$

Thus we see that ARL calculations based on the Non-Central Chi-Square distribution have incorporated assumptions of normality.

While univariate EWMAs are insensitive to the normality assumption, the MEWMA (Lowry, et. al, 1992) uses the Hotelling T^2 on simultaneous univariate EWMAs, with its accompanying normality assumptions. Departures from this assumption are considered minimal since the individual EWMAs are weighted averages, although this has not been explored for severe departures of normality in the underlying distributions of the individual EWMAs, or for higher values of smoothing constants that reduce the amount of averaging taking place by placing greater emphasis on the most recent observation.

Chapters 2 and 3 highlighted scenarios in which monitoring regression adjusted variables tends to "magnify" a shift under conditions where the usual relationship between variables is violated after an assignable cause occurs. The regression adjustment techniques published to date have assumed that the quality characteristics of interest are normally distributed with constant variance so that models fit using Ordinary Least Squares (OLS) are appropriate. Montgomery and Peck (1992) summarized problems associated with departures from these assumptions:

> Although small departures from normality do not affect the model greatly, gross nonnormality is potentially more serious as the t- or F-statistics, and confidence and prediction intervals depend on the normality assumption. Furthermore if the errors come from a distribution with thicker or heavier tails than the normal, least squares fit may be sensitive to a small subset of the data. Heavy-tailed error distributions often generate outliers that "pull" the least squares fit too much in their direction."

While data transformations (on y or x) are often used to reduce the degree to which these assumptions are violated, they do not always perform satisfactorily. Myers and Montgomery (1997) explored the analysis of two factorial experiments with non-normal response data, comparing transformation-based normal theory models to those using Generalized Linear Models (GLM) theory. In both cases, models generated using GLM theory provided substantially narrower confidence intervals around predicted values of the mean response. Furthermore, the transformation-based ordinary least squares (OLS) model for the defects response often provided negative estimates of the number of defects in an important region of interest.

The interest in this chapter is in determining if similar improvements may be obtained in regression adjustment monitoring methods through the use of GLM theory, when the usual assumptions of normality and constant variance are violated. The remainder of this chapter provides an overview of GLM theory and presents an industrial example of its application in a regression adjustment monitoring procedure.

Generalized Linear Models

Generalized Linear Models consist of three components: 1) a random component; 2) a systematic component; and 3) a link between the random and systematic components.

The random component has a distribution in the exponential family of the form (McCullagh and Nelder, 1983):

$$f_Y(y;\theta,\phi) = \exp\{(y\theta - b(\theta)) / a(\phi) + c(y,\phi)\} \tag{4-3}$$

where a(·), b(·), and c(·) are specific functions (pieces) of an original distribution that allow the original distribution to be specified in this form. These functions are such that $a(\phi) = \phi/w$ and $c = c(y, \phi/w)$ where w is a "known" weight for each observation (in parameter estimation, SAS PROC GENMOD starts with all weights = 1). The Normal, Binomial, Poisson, Gamma, and Inverse Gaussian are examples of distributions that can be expressed in the form of (4-3), and are available for use in SAS PROC GENMOD. Nelder and McCullagh's (1983) Table 2.1 contains expressions for a(·), b(·), and c(·) that allow these distributions to be expressed in exponential form (4-3). This differs from traditional linear models where the random component is assumed to be normally

distributed; however traditional linear models may be represented in this generalized form.

The systematic component is a linear predictor, η, given by:

$$\eta = \sum_{j=1}^{p} \mathbf{x}_j\beta_j \tag{4-4}$$

where $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_p$ is a set of regressor variables, and $\beta_1, \beta_2, ..., \beta_j$ are coefficients determined by Maximum Likelihood Estimation (using log likelihood functions). This systematic component is in the same form as in traditional linear models.

The link component is used to describe how the expected value of the random component is related to the linear predictor (systematic component). This may be any monotonic differentiable function that correctly relates the two components. As an example, the expected value of a binomial distribution lies between 0 and 1, so a link function should map the interval (0,1) to the whole real line (McCullagh and Nelder, 1983). In the form of expression (4-3), θ is defined as the canonical parameter. When the systematic effects are additive on the scale produced by the link, then $\theta = \eta$, and the link that accomplishes this is called the canonical link. While it is not mandatory to use a canonical link, they often make the most sense and are convenient (McCullagh and Nelder, 1983).

In classical linear models, the random component is normal, and the link function is the identity function (the expected value of the random component is linearly related to the systematic component).

Using a canonical link,

$$\theta = \eta = \sum_{j=1}^{p} \mathbf{x}_j\beta_j \ , \tag{4-5}$$

so the unknown parameters of (4-3) and (4-4) are the β_j and w_i. Estimation of these parameters is accomplished by Iteratively Re-weighted Least Squares (IRLS). This iterative procedure is used to update estimates of these parameters until an objective function called the "Deviance" is minimized. The Deviance is given by

$$D(\mathbf{y};\mu) = 2\phi[L(\mathbf{y};\mathbf{y})\text{-}L(\mu;\mathbf{y})] \tag{4-6}$$

where $L(\mathbf{y};\mathbf{y})$ is the maximum log-likelihood achievable for an unrestricted model and $L(\mu;\mathbf{y})$ is the maximum log-likelihood achievable for the parameter estimates applied to the candidate set of regressor variables specified in (restricted to) the model. Because $L(\mathbf{y};\mathbf{y})$ does not depend on the parameters, maximizing $L(\mu;\mathbf{y})$ is equivalent to minimizing $D(\mathbf{y};\mu)$ (McCullagh and Nelder, 1983).

Additional details on parameter estimation and goodness of fit testing using SAS PROC GENMOD can be found in SAS Technical Report P-243 and in Myers and Montgomery (1997). Procedures involving residual plots to detect non-linearity in Generalized Linear Models were compared in Wang (1987). Williams (1987) presented influence diagnostics based on changes in the deviance due to single case deletions. Pregibon (1980) provided tests for assessing the adequacy of the chosen link function (assuming that the error distribution and systematic component variables have been adequately selected).

The primary method used here to compare candidate models will be the difference in Deviance (normal theory models may also be estimated in the GLM framework so that Deviances may be compared). Observed standard error and ranges of the raw residuals are also compared, as are the length of prediction intervals on future observations.

Example

Consider a process in which a layer of material is deposited on a semiconductor wafer by injecting a gas into a furnace containing the wafer. Primary methods of controlling the thickness of this layer include: 1) time in the furnace; 2) temperature of the furnace; 3) gas concentration; and 4) gas flow rate.

At the conclusion of a process run, measurements are taken on two variables of interest, sheet resistivity and layer thickness. For a given material, their is usually some function of an inverse relationship between the surface area of the conducting material, and its electrical resistance. Although this is non-linear over a large range, the amount of curvature shown over a more narrow control region may be reasonably linear (though quadratic terms could be easily added to a model if required). A plot of resistivity versus thickness for the process under consideration is in Figure 4-1 (note: the four clustered points identified as "A" are far enough away from the rest of the data and the apparent slope that they are considered outliers and are ignored for model fitting purposes, but will be used to see if the resulting control method detects these points).

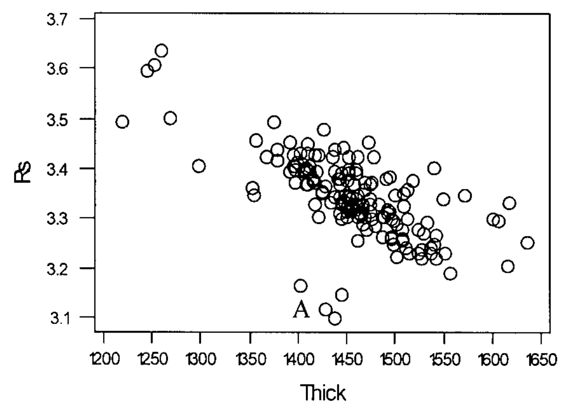


Figure 4-1. Sheet resistivity versus layer thickness.

Sheet resistivity is determined by setting the layer thickness which is controlled using factors mentioned above. If thickness alone were monitored, it is possible that contamination may change the conductance properties of the material even though the surface area remained on target. It is also possible that the thickness measure itself may not adequately reflect surface area available for conductance when there are substantial physical irregularities in the layer. Monitoring sheet resistivity alone may not provide enough information with respect to assignable cause when it is off-target. Did the "recipe" simply fail to achieve the intended thickness, or must further investigation be conducted to detect contamination?

Though these variables are both measured after a single step, they may be considered as "cascading" -- under normal operating conditions, thickness determines sheet resistivity. Considering this as well as the strong relationship between the measures ($r = -.77$), this scenario is well-suited to monitoring with regression adjusted variables as

proposed by Hawkins (1993). Following this approach, thickness would be monitored using a univariate chart, and the residuals from a regression of resistivity on thickness would be monitored with a second univariate chart. A shift in thickness will show up in the former chart, but since the physical relationship between resistivity and thickness is maintained, residuals in the latter chart will remain small. A contamination problem would not show up in the former chart when thickness is correctly maintained, but residuals in the latter chart will be larger as the contamination changes the relationship between the variables.

As previously mentioned, OLS regression requires that the model errors are independent, and normally distributed with zero mean and constant variance. Figure 4-2 is a normal probability plot of the standardized residuals from a least-squares regression of resistivity on thickness. The long "right-tail" in this plot is evidence of non-normality.

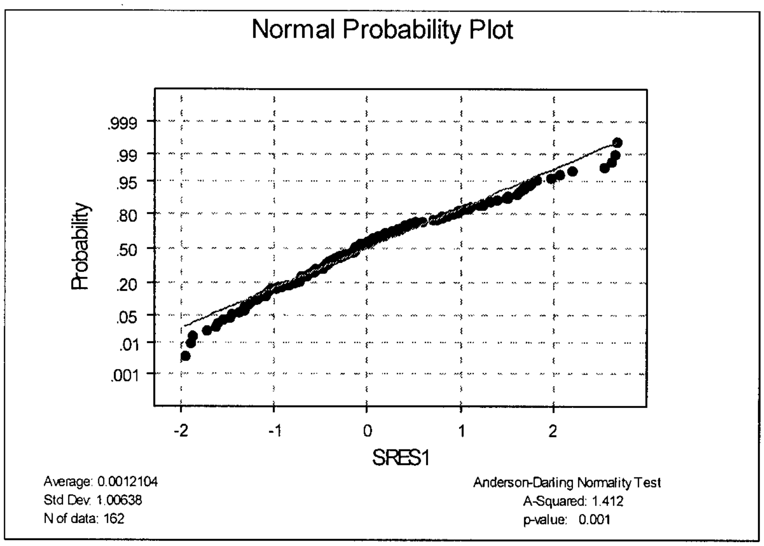


Figure 4-2. Normal probability plot of residuals from resistivity regressed on thickness.

Figure 4-3 contains a plot of the standardized residuals versus the level of the predicted value, showing a slight tendency of increased spread as the level of the predicted value increases, indicating a potential problem with the constant variance assumption (at least early in the plot).

Figure 4-4 presents a graphical autocorrelation function which suggests that the residuals are not independent.

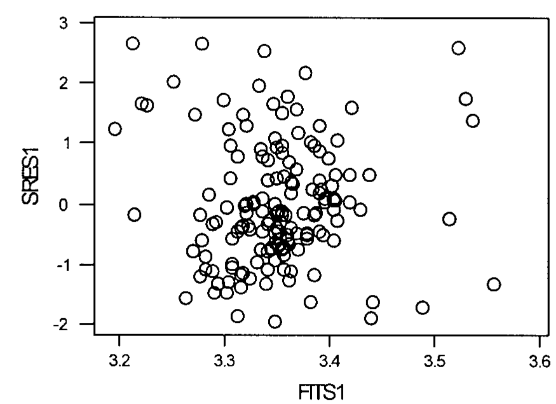
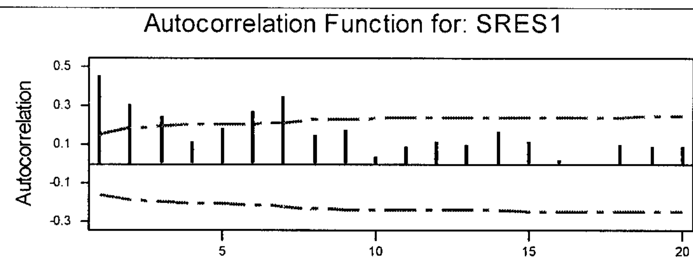


Figure 4-3. Standardized residuals versus predicted values.



Lag	Corr	t	LBQ	Lag	Corr	t	LBQ	Lag	Corr	t	LBQ
1	0.45	5.73	33.43	8	0.15	1.30	104.40	15	0.11	0.92	122.92
2	0.31	3.31	49.20	9	0.18	1.50	109.79	16	0.02	0.16	122.99
3	0.24	2.44	59.03	10	0.04	0.30	110.01	17	0.00	0.00	122.99
4	0.11	1.11	61.23	11	0.09	0.78	111.54	18	0.10	0.80	124.79
5	0.18	1.78	67.01	12	0.12	1.00	114.06	19	0.09	0.74	126.35
6	0.27	2.58	79.62	13	0.10	0.80	115.72	20	0.09	0.73	127.89
7	0.35	3.17	100.42	14	0.17	1.36	120.61				

Figure 4-4. Autocorrelation Function of standardized residuals.

While GLM theory does not rely on assumptions of normality and constant variance, it does make the assumption that observations are independent with respect to time. Violation of this assumption shown in Figure 4-4 will be initially ignored, to be discussed later in this chapter.

For comparison purposes, a suitable transformation must be selected. To reduce the skewness and stabilize variance, the natural log transformation was considered. The plots in Figures 4-5 and 4-6 show that normality and constant variance assumptions appear to be met.

Within SAS PROC GENMOD, positively skewed distributions may be modeled with the GAMMA error distribution with the inverse link being the canonical link.

The three fitted models are (first two OLS and last one GLM using GENMOD):

$$\text{Resistivity} = 4.615 - 0.000869 \times \text{Thickness} \tag{4-6}$$

$$\ln(\text{Resistivity}) = 1.583 - 0.000257 \times \text{Thickness} \tag{4-7}$$

$$\text{Resistivity} = \frac{1}{0.1870 + 0.0001 \times \text{Thickness}} \tag{4-8}$$

The positive coefficient on thickness in (4-8) may seem confusing at first -- increasing thickness usually lowers resistivity. Since the inverse link is being used, the inverse of a smaller number is larger, so (4-8) does indeed operate correctly even though the sign of the coefficient is reversed from that of the other two models.

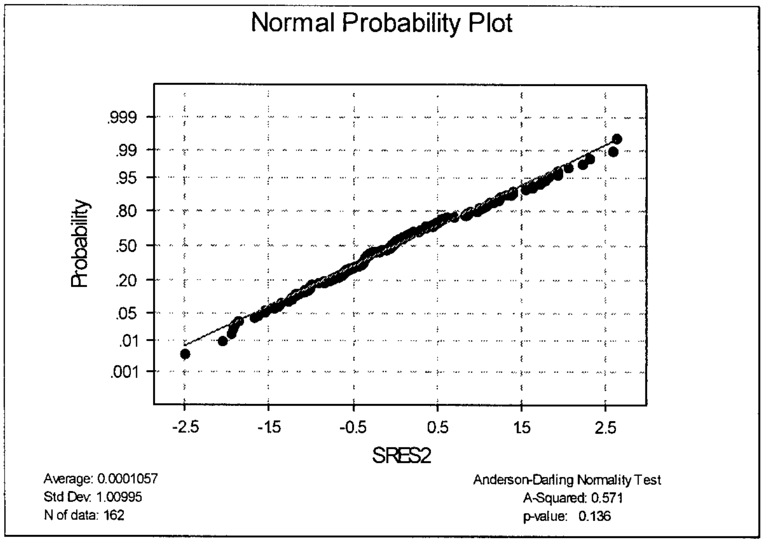


Figure 4-5. Normal probability plot of residuals from OLS regression of ln(resistivity) on thickness.

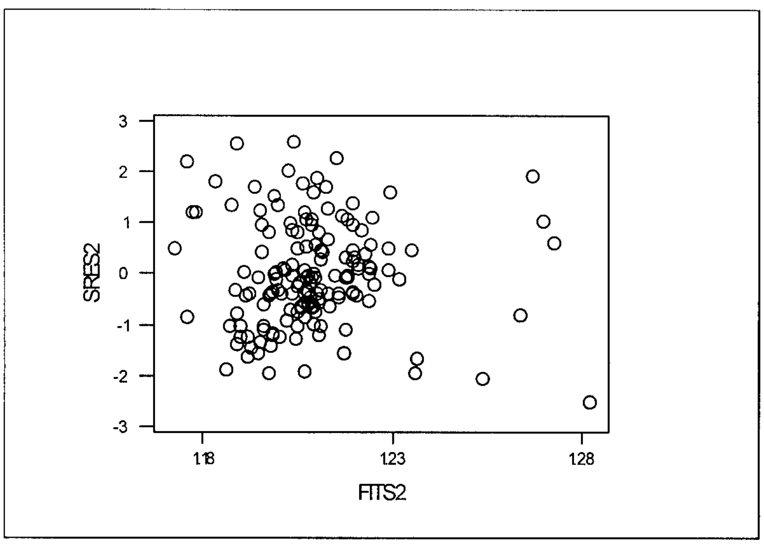


Figure 4-6. Residuals versus predicted values from OLS regression of ln(resistivity) on thickness.

The summary statistics in Table 4-1 show a small, but consistent improvement across the fitting methods. [note: the "deviance" number in Table 4-1 for the normal-theory-based models actually comes from a GENMOD fit using the normal distribution with the identity link, shown by Nelder and Wedderburn (1972) to be identical to least-squares estimation. For other comparisons to come, the least-squares procedure in PROC REG is used since the prediction intervals are more readily available]. While the improvement in Table 4-1 is admittedly small, we must recall that the departure from least-squares assumptions was not severe in the available data set.

Since the intended application is process monitoring, the prediction intervals are of considerable interest as they may be used as control limits. Myers and Montgomery (1997) show that an asymptotic, normal-theory prediction interval on a future observation (assuming the canonical link is used) is given by:

$$\hat{\mu}(x_0'\hat{\beta}) \pm z_{\alpha/2}\left[\frac{\hat{V}\mathrm{ar}(y_0)}{r(\phi)}\right]^{1/2}\sqrt{1+\hat{V}\mathrm{ar}(y_0)\left[x_0'\,\frac{(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}}{[r(\phi)]}\,x_0\right]} \tag{4-9}$$

Though not available directly in SAS PROC GENMOD, these values may be calculated from other output items available.

Table 4-1. Summary statistics for fitted models.

Measure	OLS	OLS Natural Log Transformation	GenMod Gamma Error Inverse Link
Deviance	.3561	.0314	.0313
Raw Residual SSE	.3561	.3537	.3514
Raw Residual Range	.2165	.2153	.2143
Raw Residual Std Error	.0470	.0469	.0467

For the Gamma distribution with the inverse canonical link:

$$\hat{\mu}(x_0'\hat{\beta}) = 1 / (x_0'\hat{\beta}); \tag{4-10}$$

which is the inverse of the linear predictor. The linear predictors are available using the OBSTATS option and are printed under the heading "Xbeta." The scale parameter, $r(\phi)$, is listed as the SCALE parameter under the Analysis of Parameter Estimates. For the Gamma distribution, the variance function is:

$$V(y_0) = \mu^2 = \left[\frac{1}{x_0\beta}\right]^2 \tag{4-11}$$

What remains is to calculate the term in brackets underneath the radical of eqn (4-9). With the OBSTATS option, SAS PROC GENMOD includes a quantity called "Std," which is the standard error of the linear predictor. The variance of the linear predictor is given in SAS Tech Report P-243 as:

$$\sigma_x^2 = \mathbf{x}_i'\Sigma\mathbf{x}_i \tag{4-12}$$

where Σ is the covariance of the parameter estimates, $\hat{\beta}$. Myers and Montgomery (1997) note the covariance matrix of the parameter estimates is:

$$\mathrm{Var}(\hat{\beta}) = \frac{[\mathbf{X}'\mathbf{V}\mathbf{X}]^{-1}}{[r(\phi)]^2} \tag{4-13}$$

Substituting (4-13) into (4-12) we see that squaring "Std" and multiplying by the scale parameter provides the quantity in the brackets underneath the radical in eqn (4-9).

This approach was used to form 95% prediction intervals on future observations for comparison between models [note: instead of $z_{\alpha/2}$, $t_{\alpha/2,n-p}$ was used in (4-9) to be consistent with SAS PROC REG for least-squares 95% prediction intervals -- the difference is small, but gets magnified in the links and inverse transformations being applied].

Table 4-2 contains a 95% prediction interval length comparison across the three models for the first ten observations. These points are reasonably illustrative as they contain a wide range of resistivity levels. The average prediction interval length for all 162 observations is presented at the bottom of Table 4-2. In general, the interval using the GenMod fit is slightly narrower than the others, with the transformation-based OLS intervals being slightly narrower than the OLS intervals on the raw data. An exception to this occurs at higher levels of resistivity (points 9 and 10) where the OLS intervals on the

Table 4-2. Prediction interval length comparison.

		Least Squares Fit				GenMod Fit	
		No Transformation		Ln Transformation (Untransformed)		Gamma Error Inverse Link	
Obs	Observed Value	Predicted Value	95% PI Width	Predicted Value	95% PI Width	Predicted Value	95% PI Width
1	3.302	3.2260	0.1897	3.2284	0.1821	3.2308	0.1804
2	3.298	3.2199	0.1900	3.2226	0.1818	3.2252	0.1803
3	3.361	3.3024	0.1873	3.3026	0.1840	3.3027	0.1823
4	3.402	3.2772	0.1879	3.2779	0.1829	3.2786	0.1815
5	3.205	3.2129	0.1903	3.2162	0.1817	3.2188	0.1803
6	3.347	3.2512	0.1887	3.2527	0.1825	3.2541	0.1809
7	3.335	3.2112	0.1904	3.2146	0.1816	3.2172	0.1803
8	3.252	3.1947	0.1913	3.1989	0.1817	3.2022	0.1801
9	3.496	3.5553	0.1943	3.5591	0.2054	3.5652	0.2048
10	3.408	3.4875	0.1902	3.4882	0.1971	3.4908	0.1958
Average (162 Obs)			0.1875		0.1867		0.1852

raw data are narrower.

This may seem undesirable at first; however, we must consider that the data indicated a tendency for resistivity to exhibit more variability at higher levels. It is appropriate that prediction intervals on data that exhibit non-constant variance should be a function of the variance. Figure 4-7 plots the 95% prediction interval length versus the level of the predicted resistivity (GLM predicted values) for all three models (values from the ln(resistivity) model have been untransformed). This figure shows that the GLM model and the transformation-based OLS model vary the interval length more as a function of variance at the level of the predicted value (with the GLM width slightly narrower), and that the OLS model interval length varies solely as a function of the distance from the center of the data.

This feature should be considered desirable for process monitoring of data with non-constant variance -- intervals that assume constant variance and are tighter than they

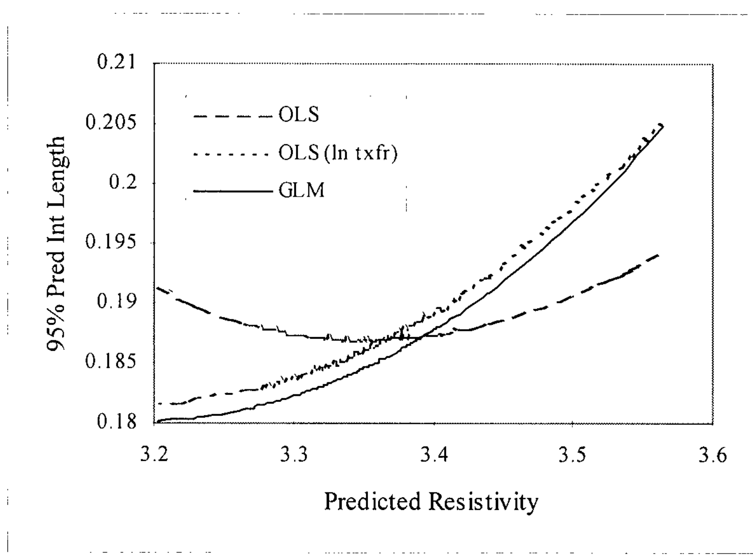


Figure 4-7. 95% prediction interval length versus predicted resistivity.

should be in regions of naturally higher variability would lead to increased incidence of false alarms.

Plots of observed values versus the prediction intervals are in Figures 4-8 and 4-9 for the transformation-based OLS and GLM fits, respectively. Both models indicate that points 4, 6, 7, 13, 34, 78 are suspect; however, even though the noted improvements in prediction intervals seem to be small, the slightly tighter interval from the GLM fit indicates an additional suspect point associated with run 66.

Point 13 in Figures 4-8 and 4-9 calls into question an assumption made when fitting the models in equations (4-6) through (4-8). Second-order terms were considered and found to be marginally statistically significant to varying degrees across the models. Since including these terms only improved the deviance by 2-3%, and because the plot in Figure 4-1 looks largely linear, the decision was made not to include the second order

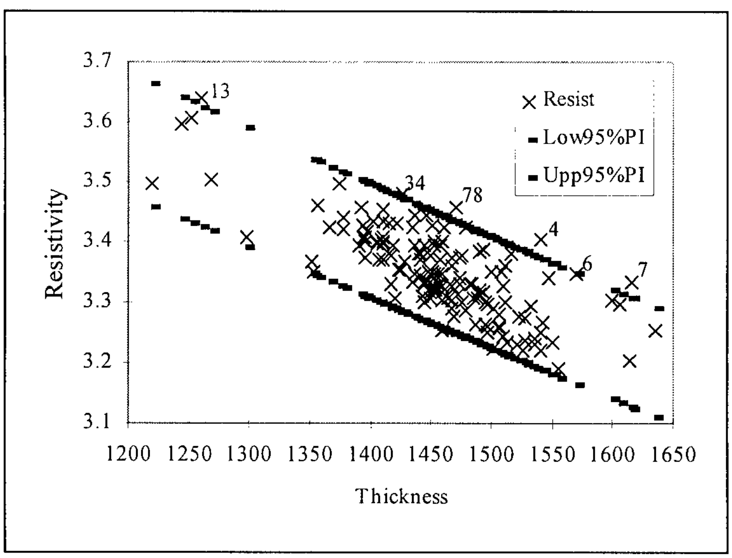


Figure 4-8. 95% Prediction intervals from OLS ln(resistivity) model.

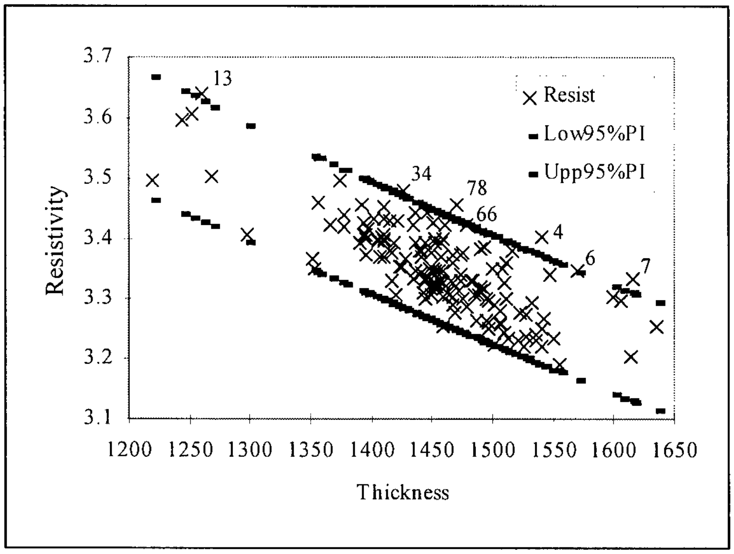


Figure 4-9. 95% Prediction interval from Generalized Linear Model.

term. Another factor supporting this decision is that multivariate statistical techniques that rely on the covariance matrix are only considering linear relationships between variables.

Knowing that the relationship between resistivity and thickness should appear more non-linear over a larger range of exploration and re-calling the marginal statistical significance of second order model terms, we should consider whether point 13 may have been "penalized" by considering only the linear relationship. This is mitigated somewhat by the natural log transformation in model (4-7) and the inverse link function in model model (4-8). Figure 4-10 shows the observed resistivity versus the 95% prediction limits based on a generalized linear model incorporating a second-order linear predictor. Points 6 and 13 are no longer outside the prediction limits. The second order model seems to work quite well over the range of the plot (the average 95% prediction interval width is reduced to 0.1836). None of the remaining points is very far outside the prediction limits.

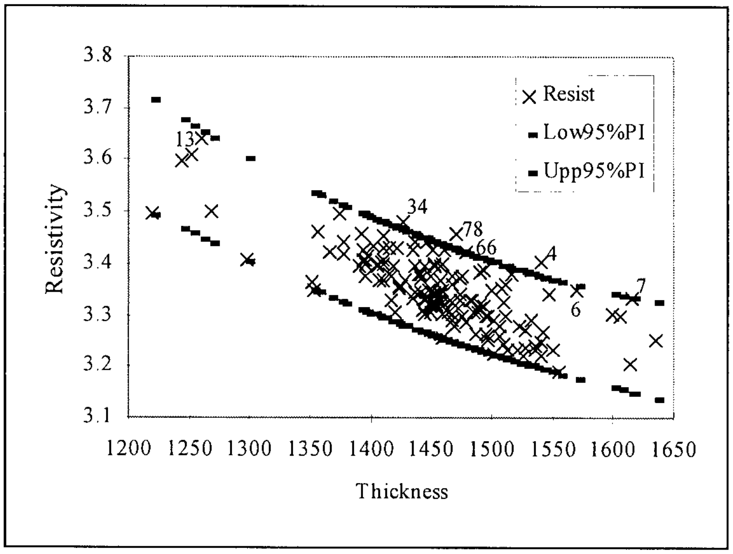


Figure 4-10. 95% prediction limits for Generalized Linear Model including second order linear predictor.

If these products were well-within specification tolerances, then the prediction intervals could be increased to higher percentages. Figure 4-11 shows 99% prediction intervals, with all points contained therein. The cluster of four points labeled A in Figure 4-1 still falls outside the prediction limits as annotated by the area A in Figure 4-11.

Trends over time are difficult to see in Figures 4-8 through 4-11. Plotting resistivity and its prediction limits in run order offer additional insight (Figure 4-12), although not as much as in traditional charting since the prediction limits vary with respect to the observed thickness [note: to reduce clutter, only the first 40 observations are shown in Figure 12]. The simultaneous application of a univariate chart on thickness, and a prediction limit chart on resistivity could be incorporated into a single graph by placing the univariate control limits for thickness as vertical lines on Figures 4-8 through 4-11.

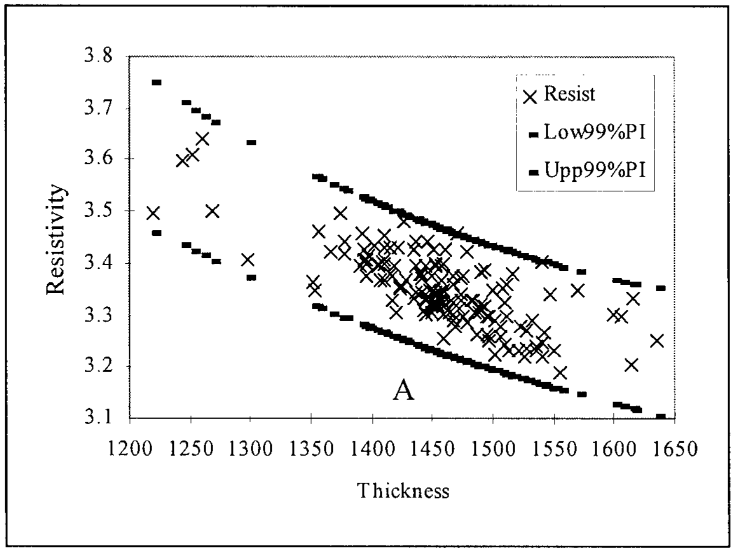


Figure 4-11. 99% prediction limits for generalized linear model including second order linear predictor.

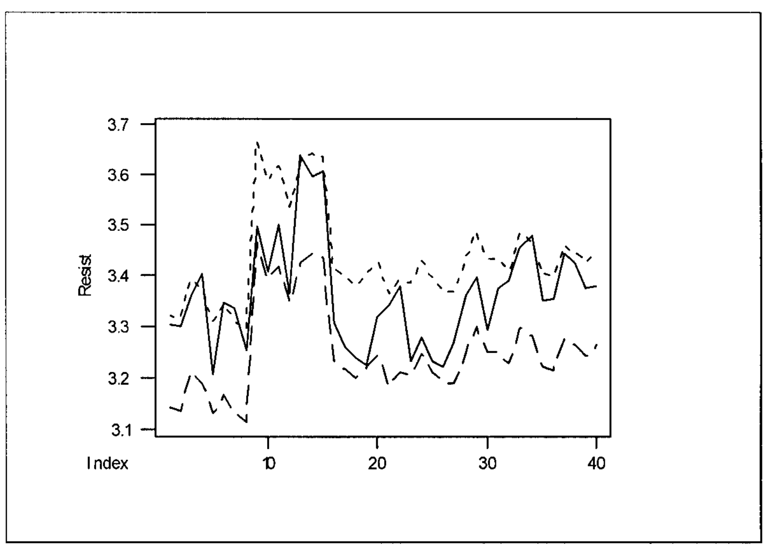


Figure 4-12. Run Order Plot (first 40 observations) of Resistivity versus 95% prediction limits from the generalized linear model.

Autocorrelation

Figure 4-4 showed the presence of moderate autocorrelation in the data. For OLS, the mean square error could be underestimated (Montgomery and Peck, 1992). While this probably has not led to wrong conclusions regarding regressor significance (only one is considered in this example scenario), it may be lead to the production of prediction limits that are misleadingly narrow. GLM also assumes independence with respect to time. When this is violated, the variance of the linear predictor would be underestimated, which in turn would lead to the production of misleadingly narrow prediction limits. This would lead to an increased false alarm rate -- analogous to results found by Harris and Ross (1991) for the univariate case.

Other Considerations

When more frequent observations are available, it may be better to give up the prediction limit charts of individual observations and use a CUSUM on averaged residuals (still not normally distributed), or an EWMA on the residuals. Despite the improved model fit using GLM, deviance residuals were still non-normal indicating a lack of asymptotic conditions shown by Pierce & Schaefer (1986). The raw residuals also remained non-normal as expected. Even so, the residual standard deviation is smaller with the GLM fit, and Hawkins has shown (assuming a reasonable relationship exists between variables) that the residual standard deviation is smaller than in the original units. Since the shift in the mean is equivalent on either scale, the shift is magnified as a

function of its residual standard error. The key remaining assumption for performance improvements from regression adjustment is that the usual relationship between variables is violated. The univariate chart on thickness would detect the situation that the residual-based chart on resistivity would not -- when thickness and resistivity move together along their normal relationship.

Summary and Conclusions

Assumptions of normality and constant variance are imbedded in many statistical monitoring procedures. While an appeal may be made to the central limit theorem so that batch means may be considered normally distributed, when process run times are long, it may not be practical to wait for several observations before obtaining a single batch mean control chart point.

For scenarios where monitoring with regression adjusted variables is appropriate, but assumptions of normality and constant variance are violated, prediction limits based on generalized linear model theory were shown to offer tighter control than OLS based methods relying on a data transformation. Though the improvements appeared small, the example is considered very conservative as the departure from these assumptions was minor, and the model was fit with process data thought to be in a reasonable state of control. Montgomery and Myers (1997) have shown larger improvements with GLM fitting under more severe departures from assumptions, with response data involving larger exploratory ranges.

Furthermore, under conditions of non-constant variance in the variable of interest, the concept that prediction limits should vary with the level of the predicted value was

emphasized, otherwise unnecessary false alarms in regions of naturally higher variability would occur. Both the transformation-based OLS and the GLM models were shown to possess this property, with the GLM model exhibiting tighter intervals. Finally, the example also demonstrated regression adjustment's flexibility to consider non-linear relationships between variables within the monitoring procedure.

APPENDIX 4.A

Supporting Data for Ordinary Least Squares Model

OLS Model Summary

OLS Fit to Resistivity/Thickness Data

Model: MODEL1
Dependent Variable: RESIST

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	0.51502	0.51502	231.374	0.0001
Error	160	0.35615	0.00223		
C Total	161	0.87117			

Root MSE	0.04718	R-square	0.5912
Dep Mean	3.35159	Adj R-sq	0.5886
C.V.	1.40768		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	4.615285	0.08316033	55.499	0.0001
THICK	1	-0.000869	0.00005712	-15.211	0.0001

Variance

Variable	DF	Inflation
INTERCEP	1	0.00000000
THICK	1	1.00000000

Durbin-Watson D	1.079
(For Number of Obs.)	162
1st Order Autocorrelation	0.450

Data, Predicted Values, and 95% Prediction Intervals

Obs	Thick	Dep Var Res	Predict Value	L95% Predict	U95% Predict	Obs	Thick	Dep Var Res	Predict Value	L95% Predict	U95% Predict
1	1599	3.302	3.2260	3.1311	3.3208	43	1436	3.345	3.3676	3.2741	3.4611
2	1606	3.298	3.2199	3.1249	3.3149	44	1375	3.497	3.4206	3.3267	3.5145
3	1511	3.361	3.3024	3.2088	3.3961	45	1392	3.456	3.4058	3.3121	3.4996
4	1540	3.402	3.2772	3.1833	3.3712	46	1461	3.425	3.3459	3.2524	3.4393
5	1614	3.205	3.2129	3.1178	3.3081	47	1446	3.444	3.3589	3.2654	3.4524
6	1570	3.347	3.2512	3.1568	3.3455	48	1428	3.368	3.3745	3.2810	3.4681
7	1616	3.335	3.2112	3.1160	3.3064	49	1435	3.425	3.3685	3.2750	3.4620
8	1635	3.252	3.1947	3.0990	3.2903	50	1458	3.394	3.3485	3.2550	3.4419
9	1220	3.496	3.5553	3.4581	3.6524	51	1460	3.347	3.3467	3.2533	3.4402
10	1298	3.408	3.4875	3.3924	3.5826	52	1452	3.400	3.3537	3.2602	3.4472
11	1269	3.502	3.5127	3.4169	3.6085	53	1421	3.430	3.3806	3.2871	3.4742
12	1352	3.365	3.4406	3.3464	3.5348	54	1436	3.395	3.3676	3.2741	3.4611
13	1260	3.640	3.5205	3.4245	3.6165	55	1528	3.273	3.2877	3.1938	3.3815
14	1244	3.598	3.5344	3.4380	3.6308	56	1482	3.329	3.3276	3.2341	3.4211
15	1252	3.608	3.5275	3.4313	3.6237	57	1493	3.317	3.3181	3.2245	3.4116
16	1489	3.305	3.3215	3.2280	3.4151	58	1451	3.350	3.3546	3.2611	3.4480
17	1506	3.258	3.3068	3.2131	3.4004	59	1457	3.370	3.3494	3.2559	3.4428
18	1527	3.237	3.2885	3.1947	3.3824	60	1442	3.371	3.3624	3.2689	3.4559
19	1501	3.223	3.3111	3.2175	3.4047	61	1410	3.432	3.3902	3.2966	3.4838
20	1473	3.315	3.3354	3.2420	3.4289	62	1456	3.345	3.3502	3.2568	3.4437
21	1548	3.341	3.2703	3.1762	3.3643	63	1474	3.340	3.3346	3.2411	3.4281
22	1517	3.379	3.2972	3.2035	3.3909	64	1448	3.350	3.3572	3.2637	3.4506
23	1522	3.231	3.2929	3.1991	3.3866	65	1491	3.313	3.3198	3.2263	3.4134
24	1469	3.277	3.3389	3.2454	3.4324	66	1478	3.424	3.3311	3.2376	3.4246
25	1513	3.232	3.3007	3.2070	3.3944	67	1416	3.431	3.3850	3.2914	3.4785
26	1541	3.221	3.2764	3.1824	3.3703	68	1402	3.434	3.3971	3.3035	3.4908
27	1542	3.268	3.2755	3.1815	3.3695	69	1491	3.320	3.3198	3.2263	3.4134
28	1468	3.359	3.3398	3.2463	3.4333	70	1444	3.330	3.3606	3.2672	3.4541
29	1406	3.396	3.3937	3.3000	3.4873	71	1486	3.305	3.3242	3.2306	3.4177
30	1467	3.290	3.3407	3.2472	3.4341	72	1474	3.330	3.3346	3.2411	3.4281
31	1468	3.375	3.3398	3.2463	3.4333	73	1442	3.380	3.3624	3.2689	3.4559
32	1494	3.387	3.3172	3.2236	3.4108	74	1459	3.318	3.3476	3.2541	3.4411
33	1410	3.452	3.3902	3.2966	3.4838	75	1459	3.400	3.3476	3.2541	3.4411
34	1426	3.480	3.3763	3.2828	3.4698	76	1506	3.280	3.3068	3.2131	3.4004
35	1500	3.350	3.3120	3.2184	3.4056	77	1467	3.326	3.3407	3.2472	3.4341
36	1508	3.351	3.3050	3.2114	3.3987	78	1471	3.457	3.3372	3.2437	3.4307
37	1437	3.442	3.3667	3.2732	3.4602	79	1495	3.300	3.3163	3.2228	3.4099
38	1451	3.426	3.3546	3.2611	3.4480	80	1497	3.297	3.3146	3.2210	3.4082
39	1476	3.376	3.3328	3.2393	3.4263	81	1444	3.394	3.3606	3.2672	3.4541
40	1449	3.378	3.3563	3.2628	3.4498	82	1483	3.329	3.3268	3.2332	3.4203
41	1452	3.394	3.3537	3.2602	3.4472	83	1420	3.305	3.3815	3.2880	3.4750
42	1509	3.325	3.3042	3.2105	3.3978	84	1464	3.308	3.3433	3.2498	3.4367

Obs	Thick	Dep Var Res	Predict Value	L95% Predict	U95% Predict	Obs	Thick	Dep Var Res	Predict Value	L95% Predict	U95% Predict
85	1458	3.342	3.3485	3.2550	3.4419	127	1407	3.370	3.3928	3.2992	3.4864
86	1453	3.327	3.3528	3.2594	3.4463	128	1479	3.285	3.3302	3.2367	3.4237
87	1440	3.380	3.3641	3.2706	3.4576	129	1395	3.407	3.4032	3.3095	3.4969
88	1475	3.299	3.3337	3.2402	3.4272	130	1354	3.350	3.4388	3.3447	3.5330
89	1453	3.340	3.3528	3.2594	3.4463	131	1394	3.407	3.4041	3.3104	3.4978
90	1444	3.301	3.3606	3.2672	3.4541	132	1495	3.263	3.3163	3.2228	3.4099
91	1417	3.329	3.3841	3.2905	3.4777	133	1434	3.335	3.3693	3.2758	3.4628
92	1396	3.375	3.4024	3.3087	3.4960	134	1523	3.278	3.2920	3.1982	3.3858
93	1367	3.424	3.4275	3.3336	3.5215	135	1510	3.243	3.3033	3.2096	3.3970
94	1501	3.291	3.3111	3.2175	3.4047	136	1450	3.305	3.3554	3.2620	3.4489
95	1406	3.399	3.3937	3.3000	3.4873	137	1452	3.319	3.3537	3.2602	3.4472
96	1394	3.428	3.4041	3.3104	3.4978	138	1556	3.191	3.2633	3.1692	3.3575
97	1410	3.369	3.3902	3.2966	3.4838	139	1487	3.265	3.3233	3.2298	3.4168
98	1391	3.395	3.4067	3.3130	3.5004	140	1474	3.372	3.3346	3.2411	3.4281
99	1445	3.332	3.3598	3.2663	3.4532	141	1496	3.260	3.3155	3.2219	3.4090
100	1490	3.383	3.3207	3.2271	3.4142	142	1550	3.232	3.2685	3.1745	3.3626
101	1460	3.325	3.3467	3.2533	3.4402	143	1532	3.292	3.2842	3.1903	3.3781
102	1424	3.352	3.3780	3.2845	3.4715	144	1456	3.330	3.3502	3.2568	3.4437
103	1424	3.355	3.3780	3.2845	3.4715	145	1466	3.330	3.3415	3.2481	3.4350
104	1452	3.346	3.3537	3.2602	3.4472	146	1512	3.300	3.3016	3.2079	3.3953
105	1425	3.355	3.3772	3.2836	3.4707	147	1458	3.330	3.3485	3.2550	3.4419
106	1446	3.336	3.3589	3.2654	3.4524	148	1463	3.311	3.3441	3.2507	3.4376
107	1411	3.398	3.3893	3.2957	3.4829	149	1449	3.317	3.3563	3.2628	3.4498
108	1455	3.321	3.3511	3.2576	3.4446	150	1536	3.231	3.2807	3.1868	3.3746
109	1442	3.344	3.3624	3.2689	3.4559	151	1526	3.220	3.2894	3.1956	3.3832
110	1415	3.379	3.3858	3.2923	3.4794	152	1455	3.319	3.3511	3.2576	3.4446
111	1398	3.416	3.4006	3.3069	3.4943	153	1443	3.310	3.3615	3.2680	3.4550
112	1415	3.380	3.3858	3.2923	3.4794	154	1452	3.320	3.3537	3.2602	3.4472
113	1417	3.376	3.3841	3.2905	3.4777	155	1506	3.261	3.3068	3.2131	3.4004
114	1396	3.399	3.4024	3.3087	3.4960	156	1460	3.255	3.3467	3.2533	3.4402
115	1378	3.420	3.4180	3.3241	3.5119	157	1450	3.322	3.3554	3.2620	3.4489
116	1357	3.460	3.4362	3.3421	3.5303	158	1540	3.249	3.2772	3.1833	3.3712
117	1410	3.370	3.3902	3.2966	3.4838	159	1460	3.303	3.3467	3.2533	3.4402
118	1396	3.407	3.4024	3.3087	3.4960	160	1497	3.250	3.3146	3.2210	3.4082
119	1378	3.441	3.4180	3.3241	3.5119	161	1536	3.240	3.2807	3.1868	3.3746
120	1410	3.399	3.3902	3.2966	3.4838	162	1468	3.303	3.3398	3.2463	3.4333
121	1418	3.395	3.3832	3.2897	3.4768						
122	1411	3.408	3.3893	3.2957	3.4829						
123	1408	3.404	3.3919	3.2983	3.4855						
124	1408	3.404	3.3919	3.2983	3.4855						
125	1408	3.404	3.3919	3.2983	3.4855						
126	1408	3.404	3.3919	3.2983	3.4855						

APPENDIX 4.B

Supporting Data For Ordinary Least Squares Model Using Natural Log Transformation

OLS (Ln Transformation) Model Summary

OLS Fit to Resistivity/Thickness Data

Model: MODEL1
Dependent Variable: LNRES

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	0.04510	0.04510	229.607	0.0001
Error	160	0.03143	0.00020		
C Total	161	0.07653			

Root MSE	0.01402	R-square	0.5893
Dep Mean	1.20920	Adj R-sq	0.5868
C.V.	1.15905		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
INTERCEP	1	1.583155	0.02470361	64.086	0.0001
THICK	1	-0.000257	0.00001697	-15.153	0.0001

Variable	DF	Variance Inflation
INTERCEP	1	0.00000000
THICK	1	1.00000000

Durbin-Watson D	1.076
(For Number of Obs.)	162
1st Order Autocorrelation	0.452

Natural Log Transformation and OLS Fit: Data, Predicted Values, and 95% Prediction Intervals

Obs	Thick	ln Thick	Dep Var Res	Dep Var ln (Res)	Predict Value	exp (Predict Value)	L95% Predict	U95% Predict	exp (L95% Predict)	exp (U95% Predict)
1	1599	7.3771	3.3019	1.1945	1.1720	3.2284	1.1438	1.2002	3.1387	3.3208
2	1606	7.3815	3.2979	1.1933	1.1702	3.2226	1.1420	1.1984	3.1330	3.3148
3	1511	7.3205	3.3609	1.2122	1.1947	3.3026	1.1668	1.2225	3.2117	3.3957
4	1540	7.3395	3.4021	1.2244	1.1872	3.2779	1.1593	1.2151	3.1877	3.3706
5	1614	7.3865	3.2050	1.1647	1.1682	3.2162	1.1399	1.1964	3.1265	3.3082
6	1570	7.3588	3.3471	1.2081	1.1795	3.2527	1.1514	1.2075	3.1626	3.3451
7	1616	7.3877	3.3351	1.2045	1.1677	3.2146	1.1394	1.1959	3.1249	3.3065
8	1635	7.3994	3.2521	1.1793	1.1628	3.1989	1.1344	1.1912	3.1093	3.2910
9	1220	7.1066	3.4959	1.2516	1.2695	3.5591	1.2406	1.2983	3.4577	3.6631
10	1298	7.1686	3.4079	1.2261	1.2494	3.4882	1.2212	1.2777	3.3913	3.5884
11	1269	7.1460	3.5019	1.2533	1.2569	3.5145	1.2284	1.2853	3.4158	3.6158
12	1352	7.2093	3.3649	1.2134	1.2355	3.4401	1.2076	1.2635	3.3454	3.5378
13	1260	7.1389	3.6401	1.2920	1.2592	3.5226	1.2307	1.2877	3.4236	3.6244
14	1244	7.1261	3.5981	1.2804	1.2633	3.5371	1.2347	1.2919	3.4373	3.6397
15	1252	7.1325	3.6082	1.2832	1.2612	3.5297	1.2327	1.2898	3.4305	3.6321
16	1489	7.3059	3.3049	1.1954	1.2003	3.3211	1.1725	1.2281	3.2301	3.4147
17	1506	7.3172	3.2580	1.1811	1.1959	3.3065	1.1681	1.2238	3.2159	3.4001
18	1527	7.3311	3.2368	1.1746	1.1905	3.2887	1.1627	1.2184	3.1986	3.3818
19	1501	7.3139	3.2230	1.1703	1.1972	3.3108	1.1694	1.2250	3.2201	3.4042
20	1473	7.2951	3.3151	1.1985	1.2044	3.3348	1.1767	1.2322	3.2437	3.4288
21	1548	7.3447	3.3411	1.2063	1.1851	3.2710	1.1572	1.2131	3.1810	3.3639
22	1517	7.3245	3.3791	1.2176	1.1931	3.2973	1.1653	1.2210	3.2069	3.3906
23	1522	7.3278	3.2310	1.1728	1.1918	3.2930	1.1640	1.2197	3.2027	3.3862
24	1469	7.2923	3.2769	1.1869	1.2054	3.3381	1.1777	1.2332	3.2469	3.4322
25	1513	7.3218	3.2320	1.1731	1.1941	3.3006	1.1663	1.2220	3.2101	3.3940
26	1541	7.3402	3.2210	1.1697	1.1869	3.2769	1.1590	1.2149	3.1867	3.3700
27	1542	7.3408	3.2681	1.1842	1.1867	3.2763	1.1588	1.2146	3.1861	3.3689
28	1468	7.2917	3.3589	1.2116	1.2057	3.3391	1.1779	1.2335	3.2475	3.4332
29	1406	7.2485	3.3960	1.2226	1.2216	3.3926	1.1938	1.2495	3.2996	3.4886
30	1467	7.2910	3.2900	1.1909	1.2060	3.3401	1.1782	1.2337	3.2485	3.4339
31	1468	7.2917	3.3750	1.2164	1.2057	3.3391	1.1779	1.2335	3.2475	3.4332
32	1494	7.3092	3.3868	1.2199	1.1990	3.3168	1.1712	1.2268	3.2259	3.4103
33	1410	7.2513	3.4522	1.2390	1.2206	3.3892	1.1928	1.2484	3.2963	3.4848
34	1426	7.2626	3.4799	1.2470	1.2165	3.3754	1.1887	1.2443	3.2828	3.4705
35	1500	7.3132	3.3501	1.2090	1.1975	3.3118	1.1697	1.2253	3.2210	3.4052
36	1508	7.3185	3.3511	1.2093	1.1954	3.3049	1.1676	1.2232	3.2143	3.3980
37	1437	7.2703	3.4422	1.2361	1.2137	3.3659	1.1859	1.2414	3.2736	3.4605
38	1451	7.2800	3.4260	1.2314	1.2101	3.3538	1.1823	1.2378	3.2619	3.4480
39	1476	7.2971	3.3760	1.2167	1.2036	3.3321	1.1759	1.2314	3.2411	3.4260
40	1449	7.2786	3.3781	1.2173	1.2106	3.3555	1.1828	1.2384	3.2635	3.4501
41	1452	7.2807	3.3940	1.2220	1.2098	3.3528	1.1821	1.2376	3.2612	3.4473
42	1509	7.3192	3.3251	1.2015	1.1952	3.3042	1.1673	1.2230	3.2133	3.3974

Obs	Thick	ln Thick	Dep Var Res	Dep Var ln (Res)	Predict Value	exp (Predict Value)	L95% Predict	U95% Predict	exp (L95% Predict)	exp (U95% Predict)
43	1436	7.2696	3.3451	1.2075	1.2139	3.3666	1.1862	1.2417	3.2746	3.4615
44	1375	7.2262	3.4970	1.2519	1.2296	3.4199	1.2017	1.2575	3.3258	3.5166
45	1392	7.2385	3.4560	1.2401	1.2252	3.4048	1.1974	1.2531	3.3115	3.5012
46	1461	7.2869	3.4250	1.2311	1.2075	3.3451	1.1797	1.2353	3.2534	3.4394
47	1446	7.2766	3.4439	1.2366	1.2114	3.3582	1.1836	1.2391	3.2661	3.4525
48	1428	7.2640	3.3679	1.2143	1.2160	3.3737	1.1882	1.2438	3.2812	3.4688
49	1435	7.2689	3.4250	1.2311	1.2142	3.3676	1.1864	1.2420	3.2753	3.4625
50	1458	7.2848	3.3940	1.2220	1.2083	3.3478	1.1805	1.2360	3.2560	3.4418
51	1460	7.2862	3.3471	1.2081	1.2078	3.3461	1.1800	1.2355	3.2544	3.4401
52	1452	7.2807	3.4001	1.2238	1.2098	3.3528	1.1821	1.2376	3.2612	3.4473
53	1421	7.2591	3.4301	1.2326	1.2178	3.3797	1.1900	1.2456	3.2871	3.4750
54	1436	7.2696	3.3950	1.2223	1.2139	3.3666	1.1862	1.2417	3.2746	3.4615
55	1528	7.3317	3.2730	1.1857	1.1903	3.2881	1.1624	1.2182	3.1976	3.3811
56	1482	7.3011	3.3291	1.2027	1.2021	3.3271	1.1743	1.2299	3.2359	3.4209
57	1493	7.3085	3.3171	1.1991	1.1993	3.3178	1.1715	1.2271	3.2268	3.4113
58	1451	7.2800	3.3501	1.2090	1.2101	3.3538	1.1823	1.2378	3.2619	3.4480
59	1457	7.2841	3.3700	1.2149	1.2085	3.3485	1.1808	1.2363	3.2570	3.4429
60	1442	7.2738	3.3710	1.2152	1.2124	3.3615	1.1846	1.2402	3.2694	3.4563
61	1410	7.2513	3.4319	1.2331	1.2206	3.3892	1.1928	1.2484	3.2963	3.4848
62	1456	7.2834	3.3451	1.2075	1.2088	3.3495	1.1810	1.2366	3.2576	3.4439
63	1474	7.2957	3.3401	1.2060	1.2042	3.3341	1.1764	1.2319	3.2427	3.4277
64	1448	7.2779	3.3501	1.2090	1.2108	3.3562	1.1831	1.2386	3.2645	3.4508
65	1491	7.3072	3.3132	1.1979	1.1998	3.3195	1.1720	1.2276	3.2284	3.4130
66	1478	7.2984	3.4240	1.2308	1.2031	3.3304	1.1754	1.2309	3.2394	3.4243
67	1416	7.2556	3.4312	1.2329	1.2191	3.3841	1.1913	1.2469	3.2914	3.4795
68	1402	7.2457	3.4339	1.2337	1.2227	3.3963	1.1949	1.2505	3.3032	3.4921
69	1491	7.3072	3.3201	1.2000	1.1998	3.3195	1.1720	1.2276	3.2284	3.4130
70	1444	7.2752	3.3301	1.2030	1.2119	3.3599	1.1841	1.2396	3.2677	3.4542
71	1486	7.3038	3.3049	1.1954	1.2011	3.3238	1.1733	1.2289	3.2326	3.4175
72	1474	7.2957	3.3301	1.2030	1.2042	3.3341	1.1764	1.2319	3.2427	3.4277
73	1442	7.2738	3.3801	1.2179	1.2124	3.3615	1.1846	1.2402	3.2694	3.4563
74	1459	7.2855	3.3181	1.1994	1.2080	3.3468	1.1803	1.2358	3.2554	3.4411
75	1459	7.2855	3.4001	1.2238	1.2080	3.3468	1.1803	1.2358	3.2554	3.4411
76	1506	7.3172	3.2799	1.1878	1.1959	3.3065	1.1681	1.2238	3.2159	3.4001
77	1467	7.2910	3.3261	1.2018	1.2060	3.3401	1.1782	1.2337	3.2485	3.4339
78	1471	7.2937	3.4570	1.2404	1.2049	3.3364	1.1772	1.2327	3.2453	3.4305
79	1495	7.3099	3.2999	1.1939	1.1988	3.3161	1.1710	1.2266	3.2252	3.4096
80	1497	7.3112	3.2970	1.1930	1.1983	3.3145	1.1704	1.2261	3.2233	3.4079
81	1444	7.2752	3.3940	1.2220	1.2119	3.3599	1.1841	1.2396	3.2677	3.4542
82	1483	7.3018	3.3291	1.2027	1.2018	3.3261	1.1741	1.2296	3.2352	3.4199
83	1420	7.2584	3.3049	1.1954	1.2180	3.3804	1.1903	1.2458	3.2881	3.4757
84	1464	7.2889	3.3079	1.1963	1.2067	3.3424	1.1790	1.2345	3.2511	3.4367

Obs	Thick	ln Thick	Dep Var Res	Dep Var ln (Res)	Predict Value	exp (Predict Value)	L95% Predict	U95% Predict	exp (L95% Predict)	exp (U95% Predict)
85	1458	7.2848	3.3421	1.2066	1.2083	3.3478	1.1805	1.2360	3.2560	3.4418
86	1453	7.2814	3.3271	1.2021	1.2096	3.3521	1.1818	1.2373	3.2602	3.4463
87	1440	7.2724	3.3801	1.2179	1.2129	3.3632	1.1851	1.2407	3.2710	3.4580
88	1475	7.2964	3.2989	1.1936	1.2039	3.3331	1.1761	1.2317	3.2417	3.4271
89	1453	7.2814	3.3401	1.2060	1.2096	3.3521	1.1818	1.2373	3.2602	3.4463
90	1444	7.2752	3.3009	1.1942	1.2119	3.3599	1.1841	1.2396	3.2677	3.4542
91	1417	7.2563	3.3291	1.2027	1.2188	3.3831	1.1910	1.2466	3.2904	3.4785
92	1396	7.2414	3.3750	1.2164	1.2242	3.4014	1.1964	1.2521	3.3082	3.4977
93	1367	7.2204	3.4240	1.2308	1.2317	3.4271	1.2038	1.2596	3.3328	3.5240
94	1501	7.3139	3.2910	1.1912	1.1972	3.3108	1.1694	1.2250	3.2201	3.4042
95	1406	7.2485	3.3991	1.2235	1.2216	3.3926	1.1938	1.2495	3.2996	3.4886
96	1394	7.2399	3.4281	1.2320	1.2247	3.4031	1.1969	1.2526	3.3098	3.4994
97	1410	7.2513	3.3689	1.2146	1.2206	3.3892	1.1928	1.2484	3.2963	3.4848
98	1391	7.2378	3.3950	1.2223	1.2255	3.4059	1.1977	1.2533	3.3125	3.5019
99	1445	7.2759	3.3321	1.2036	1.2116	3.3589	1.1839	1.2394	3.2671	3.4535
100	1490	7.3065	3.3831	1.2188	1.2001	3.3204	1.1723	1.2278	3.2294	3.4137
101	1460	7.2862	3.3251	1.2015	1.2078	3.3461	1.1800	1.2355	3.2544	3.4401
102	1424	7.2612	3.3521	1.2096	1.2170	3.3770	1.1892	1.2448	3.2845	3.4722
103	1424	7.2612	3.3552	1.2105	1.2170	3.3770	1.1892	1.2448	3.2845	3.4722
104	1452	7.2807	3.3461	1.2078	1.2098	3.3528	1.1821	1.2376	3.2612	3.4473
105	1425	7.2619	3.3552	1.2105	1.2168	3.3764	1.1890	1.2445	3.2838	3.4712
106	1446	7.2766	3.3361	1.2048	1.2114	3.3582	1.1836	1.2391	3.2661	3.4525
107	1411	7.2521	3.3980	1.2232	1.2204	3.3885	1.1926	1.2482	3.2956	3.4841
108	1455	7.2828	3.3211	1.2003	1.2090	3.3501	1.1813	1.2368	3.2586	3.4446
109	1442	7.2738	3.3441	1.2072	1.2124	3.3615	1.1846	1.2402	3.2694	3.4563
110	1415	7.2549	3.3791	1.2176	1.2193	3.3848	1.1915	1.2471	3.2920	3.4802
111	1398	7.2428	3.4161	1.2285	1.2237	3.3997	1.1959	1.2515	3.3065	3.4956
112	1415	7.2549	3.3801	1.2179	1.2193	3.3848	1.1915	1.2471	3.2920	3.4802
113	1417	7.2563	3.3760	1.2167	1.2188	3.3831	1.1910	1.2466	3.2904	3.4785
114	1396	7.2414	3.3991	1.2235	1.2242	3.4014	1.1964	1.2521	3.3082	3.4977
115	1378	7.2284	3.4199	1.2296	1.2288	3.4171	1.2010	1.2567	3.3234	3.5138
116	1357	7.2130	3.4601	1.2413	1.2342	3.4356	1.2063	1.2622	3.3411	3.5332
117	1410	7.2513	3.3700	1.2149	1.2206	3.3892	1.1928	1.2484	3.2963	3.4848
118	1396	7.2414	3.4069	1.2258	1.2242	3.4014	1.1964	1.2521	3.3082	3.4977
119	1378	7.2284	3.4411	1.2358	1.2288	3.4171	1.2010	1.2567	3.3234	3.5138
120	1410	7.2513	3.3991	1.2235	1.2206	3.3892	1.1928	1.2484	3.2963	3.4848
121	1418	7.2570	3.3950	1.2223	1.2186	3.3824	1.1908	1.2464	3.2897	3.4778
122	1411	7.2521	3.4079	1.2261	1.2204	3.3885	1.1926	1.2482	3.2956	3.4841
123	1408	7.2499	3.4042	1.2250	1.2211	3.3909	1.1933	1.2489	3.2979	3.4865
124	1408	7.2499	3.4042	1.2250	1.2211	3.3909	1.1933	1.2489	3.2979	3.4865
125	1408	7.2499	3.4042	1.2250	1.2211	3.3909	1.1933	1.2489	3.2979	3.4865
126	1408	7.2499	3.4042	1.2250	1.2211	3.3909	1.1933	1.2489	3.2979	3.4865

Obs	Thick	ln Thick	Dep Var Res	Dep Var ln (Res)	Predict Value	exp (Predict Value)	L95% Predict	U95% Predict	exp (L95% Predict)	exp (U95% Predict)
127	1407	7.2492	3.3700	1.2149	1.2214	3.3919	1.1936	1.2492	3.2989	3.4876
128	1479	7.2991	3.2851	1.1894	1.2029	3.3298	1.1751	1.2307	3.2385	3.4236
129	1395	7.2406	3.4069	1.2258	1.2245	3.4025	1.1966	1.2523	3.3088	3.4984
130	1354	7.2108	3.3501	1.2090	1.2350	3.4384	1.2071	1.2630	3.3438	3.5360
131	1394	7.2399	3.4069	1.2258	1.2247	3.4031	1.1969	1.2526	3.3098	3.4994
132	1495	7.3099	3.2628	1.1826	1.1988	3.3161	1.1710	1.2266	3.2252	3.4096
133	1434	7.2682	3.3351	1.2045	1.2144	3.3683	1.1867	1.2422	3.2763	3.4632
134	1523	7.3284	3.2779	1.1872	1.1916	3.2923	1.1637	1.2194	3.2018	3.3852
135	1510	7.3199	3.2430	1.1765	1.1949	3.3032	1.1671	1.2227	3.2127	3.3963
136	1450	7.2793	3.3049	1.1954	1.2103	3.3545	1.1826	1.2381	3.2628	3.4491
137	1452	7.2807	3.3191	1.1997	1.2098	3.3528	1.1821	1.2376	3.2612	3.4473
138	1556	7.3499	3.1909	1.1603	1.1831	3.2645	1.1551	1.2111	3.1743	3.3572
139	1487	7.3045	3.2651	1.1833	1.2008	3.3228	1.1730	1.2286	3.2317	3.4164
140	1474	7.2957	3.3720	1.2155	1.2042	3.3341	1.1764	1.2319	3.2427	3.4277
141	1496	7.3106	3.2599	1.1817	1.1985	3.3151	1.1707	1.2263	3.2242	3.4086
142	1550	7.3460	3.2320	1.1731	1.1846	3.2694	1.1567	1.2126	3.1794	3.3622
143	1532	7.3343	3.2920	1.1915	1.1893	3.2848	1.1614	1.2171	3.1944	3.3774
144	1456	7.2834	3.3301	1.2030	1.2088	3.3495	1.1810	1.2366	3.2576	3.4439
145	1466	7.2903	3.3301	1.2030	1.2062	3.3408	1.1785	1.2340	3.2495	3.4349
146	1512	7.3212	3.2999	1.1939	1.1944	3.3016	1.1666	1.2222	3.2111	3.3946
147	1458	7.2848	3.3301	1.2030	1.2083	3.3478	1.1805	1.2360	3.2560	3.4418
148	1463	7.2882	3.3112	1.1973	1.2070	3.3434	1.1792	1.2348	3.2518	3.4377
149	1449	7.2786	3.3171	1.1991	1.2106	3.3555	1.1828	1.2384	3.2635	3.4501
150	1536	7.3369	3.2310	1.1728	1.1882	3.2812	1.1603	1.2161	3.1909	3.3740
151	1526	7.3304	3.2201	1.1694	1.1908	3.2897	1.1629	1.2187	3.1992	3.3828
152	1455	7.2828	3.3191	1.1997	1.2090	3.3501	1.1813	1.2368	3.2586	3.4446
153	1443	7.2745	3.3098	1.1969	1.2121	3.3605	1.1844	1.2399	3.2687	3.4553
154	1452	7.2807	3.3201	1.2000	1.2098	3.3528	1.1821	1.2376	3.2612	3.4473
155	1506	7.3172	3.2609	1.1820	1.1959	3.3065	1.1681	1.2238	3.2159	3.4001
156	1460	7.2862	3.2550	1.1802	1.2078	3.3461	1.1800	1.2355	3.2544	3.4401
157	1450	7.2793	3.3221	1.2006	1.2103	3.3545	1.1826	1.2381	3.2628	3.4491
158	1540	7.3395	3.2488	1.1783	1.1872	3.2779	1.1593	1.2151	3.1877	3.3706
159	1460	7.2862	3.3029	1.1948	1.2078	3.3461	1.1800	1.2355	3.2544	3.4401
160	1497	7.3112	3.2501	1.1787	1.1983	3.3145	1.1704	1.2261	3.2233	3.4079
161	1536	7.3369	3.2401	1.1756	1.1882	3.2812	1.1603	1.2161	3.1909	3.3740
162	1468	7.2917	3.3029	1.1948	1.2057	3.3391	1.1779	1.2335	3.2475	3.4332

APPENDIX 4.C.

Supporting Data For GLM First-Order Model

GLM First-order Model Summary

GLM Fit to Resistivity/Thickness Data

The GENMOD Procedure

Model Information

Description	Value
Data Set	WORK.PHOS
Distribution	GAMMA
Link Function	POWER(-1)
Dependent Variable	RS
Observations Used	162

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	160	0.0313	0.0002
Scaled Deviance	160	161.9677	1.0123
Pearson Chi-Square	160	0.0314	0.0002
Scaled Pearson X2	160	162.6802	1.0168
Log Likelihood	.	266.9343	.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	0.1870	0.0072	669.8431	0.0001
THICK	1	0.0001	0.0000	237.4607	0.0001
SCALE	1	5174.9827	574.9796	.	.

NOTE: The scale parameter was estimated by maximum likelihood.

LR Statistics For Type 1 Analysis

Source	Deviance	DF	ChiSquare	Pr>Chi
INTERCEPT	0.0769	0	.	.
THICK	0.0313	1	145.5487	0.0001

GLM First-order Fit
Data, Predicted Values, and 95% Prediction Intervals

Obs	Thick	Resist	Pred	Xbeta	Std	Pred Error Half-Width	L95% Predict	U95% Predict
1	1599	3.302	3.2308	0.3095	0.000799	0.0902	3.1408	3.3212
2	1606	3.298	3.2252	0.3101	0.000831	0.0902	3.1346	3.3149
3	1511	3.361	3.3027	0.3028	0.000438	0.0912	3.2114	3.3937
4	1540	3.402	3.2786	0.3050	0.000545	0.0908	3.1879	3.3694
5	1614	3.205	3.2188	0.3107	0.000868	0.0901	3.1284	3.3087
6	1570	3.347	3.2541	0.3073	0.000670	0.0904	3.1637	3.3446
7	1616	3.335	3.2172	0.3108	0.000877	0.0901	3.1274	3.3076
8	1635	3.252	3.2022	0.3123	0.000965	0.0901	3.1120	3.2921
9	1220	3.496	3.5652	0.2805	0.001200	0.1024	3.4627	3.6675
10	1298	3.408	3.4908	0.2865	0.000833	0.0979	3.3925	3.5883
11	1269	3.502	3.5181	0.2842	0.000967	0.0994	3.4192	3.6181
12	1352	3.365	3.4411	0.2906	0.000595	0.0955	3.3457	3.5366
13	1260	3.640	3.5266	0.2836	0.001010	0.0999	3.4262	3.6260
14	1244	3.598	3.5420	0.2823	0.001085	0.1009	3.4414	3.6432
15	1252	3.608	3.5343	0.2829	0.001047	0.1004	3.4344	3.6352
16	1489	3.305	3.3212	0.3011	0.000374	0.0915	3.2296	3.4127
17	1506	3.258	3.3069	0.3024	0.000421	0.0912	3.2156	3.3981
18	1527	3.237	3.2894	0.3040	0.000494	0.0909	3.1986	3.3804
19	1501	3.223	3.3111	0.3020	0.000406	0.0913	3.2199	3.4026
20	1473	3.315	3.3347	0.2999	0.000342	0.0918	3.2426	3.4263
21	1548	3.341	3.2721	0.3056	0.000577	0.0907	3.1816	3.3629
22	1517	3.379	3.2977	0.3032	0.000458	0.0911	3.2071	3.3892
23	1522	3.231	3.2935	0.3036	0.000476	0.0910	3.2028	3.3848
24	1469	3.277	3.3382	0.2996	0.000336	0.0919	3.2459	3.4297
25	1513	3.232	3.3010	0.3029	0.000444	0.0911	3.2103	3.3926
26	1541	3.221	3.2778	0.3051	0.000549	0.0907	3.1869	3.3683
27	1542	3.268	3.2770	0.3052	0.000553	0.0907	3.1858	3.3673
28	1468	3.359	3.3390	0.2995	0.000335	0.0920	3.2469	3.4309
29	1406	3.396	3.3928	0.2947	0.000399	0.0936	3.2997	3.4869
30	1467	3.290	3.3399	0.2994	0.000334	0.0920	3.2480	3.4320
31	1468	3.375	3.3390	0.2995	0.000335	0.0920	3.2469	3.4309
32	1494	3.387	3.3170	0.3015	0.000386	0.0914	3.2253	3.4082
33	1410	3.452	3.3893	0.2950	0.000388	0.0935	3.2964	3.4833
34	1426	3.480	3.3753	0.2963	0.000351	0.0930	3.2820	3.4679
35	1500	3.350	3.3119	0.3019	0.000403	0.0914	3.2210	3.4037
36	1508	3.351	3.3052	0.3026	0.000428	0.0912	3.2135	3.3959
37	1437	3.442	3.3657	0.2971	0.000334	0.0927	3.2732	3.4586
38	1451	3.426	3.3536	0.2982	0.000326	0.0923	3.2611	3.4458
39	1476	3.376	3.3322	0.3001	0.000347	0.0918	3.2404	3.4240
40	1449	3.378	3.3553	0.2980	0.000326	0.0924	3.2633	3.4481
41	1452	3.394	3.3527	0.2983	0.000326	0.0923	3.2600	3.4446
42	1509	3.325	3.3044	0.3026	0.000431	0.0912	3.2135	3.3959

Obs	Thick	Resist	Pred	Xbeta	Std	Pred Error Half-Width	L95% Predict	U95% Predict
43	1436	3.345	3.3666	0.2970	0.000336	0.0927	3.2743	3.4597
44	1375	3.497	3.4204	0.2924	0.000504	0.0946	3.3254	3.5146
45	1392	3.456	3.4052	0.2937	0.000442	0.0940	3.3108	3.4989
46	1461	3.425	3.3450	0.2990	0.000329	0.0921	3.2524	3.4366
47	1446	3.444	3.3579	0.2978	0.000327	0.0925	3.2655	3.4504
48	1428	3.368	3.3735	0.2964	0.000347	0.0929	3.2809	3.4668
49	1435	3.425	3.3674	0.2970	0.000337	0.0927	3.2743	3.4597
50	1458	3.394	3.3476	0.2987	0.000327	0.0922	3.2556	3.4400
51	1460	3.347	3.3459	0.2989	0.000328	0.0921	3.2535	3.4377
52	1452	3.400	3.3527	0.2983	0.000326	0.0923	3.2600	3.4446
53	1421	3.430	3.3796	0.2959	0.000361	0.0931	3.2864	3.4727
54	1436	3.395	3.3666	0.2970	0.000336	0.0927	3.2743	3.4597
55	1528	3.273	3.2885	0.3041	0.000498	0.0909	3.1975	3.3793
56	1482	3.329	3.3271	0.3006	0.000358	0.0917	3.2350	3.4183
57	1493	3.317	3.3178	0.3014	0.000384	0.0915	3.2264	3.4093
58	1451	3.350	3.3536	0.2982	0.000326	0.0923	3.2611	3.4458
59	1457	3.370	3.3484	0.2986	0.000327	0.0922	3.2567	3.4412
60	1442	3.371	3.3614	0.2975	0.000330	0.0926	3.2688	3.4539
61	1410	3.432	3.3893	0.2950	0.000388	0.0935	3.2964	3.4833
62	1456	3.345	3.3493	0.2986	0.000326	0.0922	3.2567	3.4412
63	1474	3.340	3.3339	0.2999	0.000343	0.0919	3.2426	3.4263
64	1448	3.350	3.3562	0.2980	0.000326	0.0924	3.2633	3.4481
65	1491	3.313	3.3195	0.3013	0.000379	0.0915	3.2275	3.4104
66	1478	3.424	3.3305	0.3003	0.000350	0.0917	3.2383	3.4217
67	1416	3.431	3.3840	0.2955	0.000372	0.0933	3.2908	3.4774
68	1402	3.434	3.3964	0.2944	0.000410	0.0937	3.3030	3.4905
69	1491	3.320	3.3195	0.3013	0.000379	0.0915	3.2275	3.4104
70	1444	3.330	3.3596	0.2977	0.000328	0.0925	3.2666	3.4516
71	1486	3.305	3.3237	0.3009	0.000367	0.0916	3.2318	3.4150
72	1474	3.330	3.3339	0.2999	0.000343	0.0919	3.2426	3.4263
73	1442	3.380	3.3614	0.2975	0.000330	0.0926	3.2688	3.4539
74	1459	3.318	3.3467	0.2988	0.000328	0.0922	3.2546	3.4389
75	1459	3.400	3.3467	0.2988	0.000328	0.0922	3.2546	3.4389
76	1506	3.280	3.3069	0.3024	0.000421	0.0912	3.2156	3.3981
77	1467	3.326	3.3399	0.2994	0.000334	0.0920	3.2480	3.4320
78	1471	3.457	3.3365	0.2997	0.000339	0.0919	3.2448	3.4286
79	1495	3.300	3.3161	0.3016	0.000389	0.0914	3.2242	3.4071
80	1497	3.297	3.3144	0.3017	0.000395	0.0914	3.2232	3.4059
81	1444	3.394	3.3596	0.2977	0.000328	0.0925	3.2666	3.4516
82	1483	3.329	3.3263	0.3006	0.000360	0.0917	3.2350	3.4183
83	1420	3.305	3.3805	0.2958	0.000363	0.0932	3.2875	3.4738
84	1464	3.308	3.3424	0.2992	0.000331	0.0920	3.2502	3.4343

Obs	Thick	Resist	Pred	Xbeta	Std	Pred Error Half-Width	L95% Predict	U95% Predict
85	1458	3.342	3.3476	0.2987	0.000327	0.0922	3.2556	3.4400
86	1453	3.327	3.3519	0.2983	0.000326	0.0923	3.2600	3.4446
87	1440	3.380	3.3631	0.2973	0.000331	0.0926	3.2710	3.4562
88	1475	3.299	3.3330	0.3000	0.000345	0.0918	3.2415	3.4252
89	1453	3.340	3.3519	0.2983	0.000326	0.0923	3.2600	3.4446
90	1444	3.301	3.3596	0.2977	0.000328	0.0925	3.2666	3.4516
91	1417	3.329	3.3831	0.2956	0.000370	0.0932	3.2897	3.4762
92	1396	3.375	3.4017	0.2940	0.000429	0.0939	3.3075	3.4953
93	1367	3.424	3.4276	0.2918	0.000534	0.0949	3.3321	3.5219
94	1501	3.291	3.3111	0.3020	0.000406	0.0913	3.2199	3.4026
95	1406	3.399	3.3928	0.2947	0.000399	0.0936	3.2997	3.4869
96	1394	3.428	3.4034	0.2938	0.000436	0.0940	3.3097	3.4976
97	1410	3.369	3.3893	0.2950	0.000388	0.0935	3.2964	3.4833
98	1391	3.395	3.4061	0.2936	0.000446	0.0941	3.3119	3.5001
99	1445	3.332	3.3588	0.2977	0.000328	0.0925	3.2666	3.4516
100	1490	3.383	3.3203	0.3012	0.000376	0.0915	3.2285	3.4116
101	1460	3.325	3.3459	0.2989	0.000328	0.0921	3.2535	3.4377
102	1424	3.352	3.3770	0.2961	0.000355	0.0931	3.2842	3.4703
103	1424	3.355	3.3770	0.2961	0.000355	0.0931	3.2842	3.4703
104	1452	3.346	3.3527	0.2983	0.000326	0.0923	3.2600	3.4446
105	1425	3.355	3.3761	0.2962	0.000353	0.0930	3.2831	3.4691
106	1446	3.336	3.3579	0.2978	0.000327	0.0925	3.2655	3.4504
107	1411	3.398	3.3884	0.2951	0.000385	0.0934	3.2952	3.4821
108	1455	3.321	3.3502	0.2985	0.000326	0.0923	3.2578	3.4423
109	1442	3.344	3.3614	0.2975	0.000330	0.0926	3.2688	3.4539
110	1415	3.379	3.3849	0.2954	0.000375	0.0933	3.2919	3.4786
111	1398	3.416	3.3999	0.2941	0.000423	0.0938	3.3064	3.4940
112	1415	3.380	3.3849	0.2954	0.000375	0.0933	3.2919	3.4786
113	1417	3.376	3.3831	0.2956	0.000370	0.0932	3.2897	3.4762
114	1396	3.399	3.4017	0.2940	0.000429	0.0939	3.3075	3.4953
115	1378	3.420	3.4177	0.2926	0.000492	0.0945	3.3231	3.5121
116	1357	3.460	3.4366	0.2910	0.000575	0.0953	3.3411	3.5317
117	1410	3.370	3.3893	0.2950	0.000388	0.0935	3.2964	3.4833
118	1396	3.407	3.4017	0.2940	0.000429	0.0939	3.3075	3.4953
119	1378	3.441	3.4177	0.2926	0.000492	0.0945	3.3231	3.5121
120	1410	3.399	3.3893	0.2950	0.000388	0.0935	3.2964	3.4833
121	1418	3.395	3.3823	0.2957	0.000368	0.0932	3.2886	3.4750
122	1411	3.408	3.3884	0.2951	0.000385	0.0934	3.2952	3.4821
123	1408	3.404	3.3911	0.2949	0.000393	0.0935	3.2975	3.4845
124	1408	3.404	3.3911	0.2949	0.000393	0.0935	3.2975	3.4845
125	1408	3.404	3.3911	0.2949	0.000393	0.0935	3.2975	3.4845
126	1408	3.404	3.3911	0.2949	0.000393	0.0935	3.2975	3.4845

Obs	Thick	Resist	Pred	Xbeta	Std	Pred Error Half-Width	L95% Predict	U95% Predict
127	1407	3.370	3.3919	0.2948	0.000396	0.0936	3.2986	3.4857
128	1479	3.285	3.3296	0.3003	0.000352	0.0917	3.2383	3.4217
129	1395	3.407	3.4025	0.2939	0.000433	0.0939	3.3086	3.4965
130	1354	3.350	3.4393	0.2908	0.000587	0.0954	3.3434	3.5342
131	1394	3.407	3.4034	0.2938	0.000436	0.0940	3.3097	3.4976
132	1495	3.263	3.3161	0.3016	0.000389	0.0914	3.2242	3.4071
133	1434	3.335	3.3683	0.2969	0.000338	0.0928	3.2754	3.4609
134	1523	3.278	3.2927	0.3037	0.000480	0.0910	3.2017	3.3837
135	1510	3.243	3.3035	0.3027	0.000434	0.0912	3.2124	3.3948
136	1450	3.305	3.3545	0.2981	0.000326	0.0924	3.2622	3.4470
137	1452	3.319	3.3527	0.2983	0.000326	0.0923	3.2600	3.4446
138	1556	3.191	3.2655	0.3062	0.000610	0.0906	3.1753	3.3564
139	1487	3.265	3.3229	0.3009	0.000369	0.0916	3.2318	3.4150
140	1474	3.372	3.3339	0.2999	0.000343	0.0919	3.2426	3.4263
141	1496	3.260	3.3153	0.3016	0.000392	0.0914	3.2242	3.4071
142	1550	3.232	3.2704	0.3058	0.000585	0.0906	3.1795	3.3607
143	1532	3.292	3.2852	0.3044	0.000513	0.0908	3.1943	3.3760
144	1456	3.330	3.3493	0.2986	0.000326	0.0922	3.2567	3.4412
145	1466	3.330	3.3407	0.2993	0.000333	0.0920	3.2491	3.4331
146	1512	3.300	3.3019	0.3029	0.000441	0.0911	3.2103	3.3925
147	1458	3.330	3.3476	0.2987	0.000327	0.0922	3.2556	3.4400
148	1463	3.311	3.3433	0.2991	0.000330	0.0921	3.2513	3.4354
149	1449	3.317	3.3553	0.2980	0.000326	0.0924	3.2633	3.4481
150	1536	3.231	3.2819	0.3047	0.000529	0.0908	3.1911	3.3727
151	1526	3.220	3.2902	0.3039	0.000491	0.0909	3.1996	3.3815
152	1455	3.319	3.3502	0.2985	0.000326	0.0923	3.2578	3.4423
153	1443	3.310	3.3605	0.2976	0.000329	0.0925	3.2677	3.4528
154	1452	3.320	3.3527	0.2983	0.000326	0.0923	3.2600	3.4446
155	1506	3.261	3.3069	0.3024	0.000421	0.0912	3.2156	3.3981
156	1460	3.255	3.3459	0.2989	0.000328	0.0921	3.2535	3.4377
157	1450	3.322	3.3545	0.2981	0.000326	0.0924	3.2622	3.4470
158	1540	3.249	3.2786	0.3050	0.000545	0.0908	3.1879	3.3694
159	1460	3.303	3.3459	0.2989	0.000328	0.0921	3.2535	3.4377
160	1497	3.250	3.3144	0.3017	0.000395	0.0914	3.2232	3.4059
161	1536	3.240	3.2819	0.3047	0.000529	0.0908	3.1911	3.3727
162	1468	3.303	3.3390	0.2995	0.000335	0.0920	3.2469	3.4309

APPENDIX 4.D.

Supporting Data For GLM Second-Order Model

GLM Second-order Model Summary

GLM Fit to Resistivity/Thickness Data

The GENMOD Procedure

Model Information

Description	Value
Data Set	WORK.PHOS
Distribution	GAMMA
Link Function	POWER(-1)
Dependent Variable	RS
Observations Used	162

Criteria For Assessing Goodness Of Fit

Criterion	DF	Value	Value/DF
Deviance	159	0.0306	0.0002
Scaled Deviance	159	161.9652	1.0186
Pearson Chi-Square	159	0.0307	0.0002
Scaled Pearson X2	159	162.3942	1.0213
Log Likelihood	.	268.7482	.

Analysis Of Parameter Estimates

Parameter	DF	Estimate	Std Err	ChiSquare	Pr>Chi
INTERCEPT	1	0.0354	0.0781	0.2053	0.6505
THICK	1	0.0003	0.0001	7.0378	0.0080
THICK2	1	-0.0000	0.0000	3.8063	0.0511
SCALE	1	5292.0973	587.9959	.	.

NOTE: The scale parameter was estimated by maximum likelihood.

LR Statistics For Type 1 Analysis

Source	Deviance	DF	ChiSquare	Pr>Chi
INTERCEPT	0.0769	0	.	.
THICK	0.0313	1	145.5487	0.0001
THICK2	0.0306	1	3.6280	0.0568

GLM Second-order Fit
Data, Predicted Values, and 95% Prediction Intervals

Obs	Thick	Resist	Pred	Xbeta	Std	95%PI Half-Width	L95% Predict	U95% Predict
1	1599	3.302	3.2485	0.3078	0.00117	0.0915	3.1573	3.3404
2	1606	3.298	3.2447	0.3082	0.00126	0.0919	3.1527	3.3365
3	1511	3.361	3.3038	0.3027	0.00044	0.0902	3.2134	3.3938
4	1540	3.402	3.2840	0.3045	0.00060	0.0901	3.1940	3.3741
5	1614	3.205	3.2405	0.3086	0.00137	0.0924	3.1480	3.3329
6	1570	3.347	3.2652	0.3063	0.00085	0.0904	3.1744	3.3552
7	1616	3.335	3.2395	0.3087	0.00140	0.0926	3.1468	3.3320
8	1635	3.252	3.2299	0.3096	0.00167	0.0942	3.1358	3.3242
9	1220	3.496	3.6038	0.2775	0.00194	0.1098	3.4938	3.7134
10	1298	3.408	3.5030	0.2855	0.00097	0.0979	3.4047	3.6006
11	1269	3.502	3.5385	0.2826	0.00127	0.1011	3.4375	3.6396
12	1352	3.365	3.4428	0.2905	0.00059	0.0945	3.3479	3.5368
13	1260	3.640	3.5500	0.2817	0.00138	0.1023	3.4476	3.6522
14	1244	3.598	3.5710	0.2800	0.00159	0.1049	3.4665	3.6763
15	1252	3.608	3.5604	0.2809	0.00148	0.1035	3.4565	3.6635
16	1489	3.305	3.3199	0.3012	0.00037	0.0905	3.2296	3.4106
17	1506	3.258	3.3074	0.3024	0.00042	0.0902	3.2167	3.3971
18	1527	3.237	3.2927	0.3037	0.00051	0.0901	3.2027	3.3828
19	1501	3.223	3.3110	0.3020	0.00040	0.0903	3.2209	3.4016
20	1473	3.315	3.3322	0.3001	0.00036	0.0908	3.2414	3.4230
21	1548	3.341	3.2789	0.3050	0.00066	0.0901	3.1886	3.3688
22	1517	3.379	3.2996	0.3031	0.00046	0.0901	3.2091	3.3894
23	1522	3.231	3.2961	0.3034	0.00049	0.0901	3.2059	3.3861
24	1469	3.277	3.3354	0.2998	0.00036	0.0909	3.2447	3.4264
25	1513	3.232	3.3024	0.3028	0.00044	0.0902	3.2123	3.3927
26	1541	3.221	3.2834	0.3046	0.00060	0.0900	3.1929	3.3730
27	1542	3.268	3.2827	0.3046	0.00061	0.0901	3.1929	3.3731
28	1468	3.359	3.3362	0.2997	0.00036	0.0909	3.2457	3.4276
29	1406	3.396	3.3894	0.2950	0.00042	0.0925	3.2973	3.4824
30	1467	3.290	3.3370	0.2997	0.00036	0.0909	3.2457	3.4276
31	1468	3.375	3.3362	0.2997	0.00036	0.0909	3.2457	3.4276
32	1494	3.387	3.3162	0.3016	0.00038	0.0904	3.2253	3.4060
33	1410	3.452	3.3858	0.2954	0.00041	0.0924	3.2929	3.4776
34	1426	3.480	3.3714	0.2966	0.00039	0.0919	3.2796	3.4635
35	1500	3.350	3.3118	0.3020	0.00040	0.0903	3.2210	3.4016
36	1508	3.351	3.3060	0.3025	0.00042	0.0902	3.2156	3.3960
37	1437	3.442	3.3618	0.2975	0.00038	0.0916	3.2697	3.4530
38	1451	3.426	3.3500	0.2985	0.00036	0.0913	3.2588	3.4414
39	1476	3.376	3.3299	0.3003	0.00036	0.0907	3.2393	3.4207
40	1449	3.378	3.3516	0.2984	0.00036	0.0913	3.2599	3.4425
41	1452	3.394	3.3491	0.2986	0.00036	0.0913	3.2577	3.4402
42	1509	3.325	3.3053	0.3025	0.00043	0.0902	3.2156	3.3960

Obs	Thick	Resist	Pred	Xbeta	Std	95%PI Half-Width	L95% Predict	U95% Predict
43	1436	3.345	3.3627	0.2974	0.00038	0.0917	3.2708	3.4541
44	1375	3.497	3.4192	0.2925	0.00050	0.0935	3.3253	3.5123
45	1392	3.456	3.4026	0.2939	0.00045	0.0929	3.3096	3.4955
46	1461	3.425	3.3418	0.2992	0.00036	0.0911	3.2512	3.4333
47	1446	3.444	3.3541	0.2981	0.00037	0.0914	3.2631	3.4460
48	1428	3.368	3.3696	0.2968	0.00039	0.0919	3.2774	3.4612
49	1435	3.425	3.3635	0.2973	0.00038	0.0917	3.2719	3.4553
50	1458	3.394	3.3442	0.2990	0.00036	0.0911	3.2533	3.4356
51	1460	3.347	3.3426	0.2992	0.00036	0.0911	3.2512	3.4333
52	1452	3.400	3.3491	0.2986	0.00036	0.0913	3.2577	3.4402
53	1421	3.430	3.3758	0.2962	0.00040	0.0921	3.2840	3.4682
54	1436	3.395	3.3627	0.2974	0.00038	0.0917	3.2708	3.4541
55	1528	3.273	3.2920	0.3038	0.00052	0.0900	3.2016	3.3817
56	1482	3.329	3.3252	0.3007	0.00036	0.0906	3.2349	3.4162
57	1493	3.317	3.3169	0.3015	0.00038	0.0904	3.2263	3.4072
58	1451	3.350	3.3500	0.2985	0.00036	0.0913	3.2588	3.4414
59	1457	3.370	3.3450	0.2990	0.00036	0.0911	3.2533	3.4356
60	1442	3.371	3.3575	0.2978	0.00037	0.0915	3.2664	3.4495
61	1410	3.432	3.3858	0.2954	0.00041	0.0924	3.2929	3.4776
62	1456	3.345	3.3458	0.2989	0.00036	0.0912	3.2544	3.4368
63	1474	3.340	3.3314	0.3002	0.00036	0.0908	3.2403	3.4219
64	1448	3.350	3.3525	0.2983	0.00037	0.0914	3.2610	3.4437
65	1491	3.313	3.3184	0.3013	0.00038	0.0905	3.2285	3.4094
66	1478	3.424	3.3283	0.3005	0.00036	0.0907	3.2371	3.4185
67	1416	3.431	3.3803	0.2958	0.00040	0.0922	3.2884	3.4729
68	1402	3.434	3.3932	0.2947	0.00043	0.0926	3.3006	3.4859
69	1491	3.320	3.3184	0.3013	0.00038	0.0905	3.2285	3.4094
70	1444	3.330	3.3558	0.2980	0.00037	0.0915	3.2642	3.4472
71	1486	3.305	3.3222	0.3010	0.00037	0.0905	3.2317	3.4128
72	1474	3.330	3.3314	0.3002	0.00036	0.0908	3.2403	3.4219
73	1442	3.380	3.3575	0.2978	0.00037	0.0915	3.2664	3.4495
74	1459	3.318	3.3434	0.2991	0.00036	0.0911	3.2523	3.4345
75	1459	3.400	3.3434	0.2991	0.00036	0.0911	3.2523	3.4345
76	1506	3.280	3.3074	0.3024	0.00042	0.0902	3.2167	3.3971
77	1467	3.326	3.3370	0.2997	0.00036	0.0909	3.2457	3.4276
78	1471	3.457	3.3338	0.3000	0.00036	0.0908	3.2425	3.4242
79	1495	3.300	3.3154	0.3016	0.00039	0.0904	3.2252	3.4061
80	1497	3.297	3.3140	0.3018	0.00039	0.0903	3.2231	3.4038
81	1444	3.394	3.3558	0.2980	0.00037	0.0915	3.2642	3.4472
82	1483	3.329	3.3245	0.3008	0.00037	0.0906	3.2339	3.4151
83	1420	3.305	3.3767	0.2961	0.00040	0.0921	3.2851	3.4694
84	1464	3.308	3.3394	0.2995	0.00036	0.0910	3.2479	3.4299

Obs	Thick	Resist	Pred	Xbeta	Std	95%PI Half-Width	L95% Predict	U95% Predict
85	1458	3.342	3.3442	0.2990	0.00036	0.0911	3.2533	3.4356
86	1453	3.327	3.3483	0.2987	0.00036	0.0912	3.2566	3.4391
87	1440	3.380	3.3592	0.2977	0.00037	0.0916	3.2675	3.4507
88	1475	3.299	3.3306	0.3002	0.00036	0.0908	3.2403	3.4219
89	1453	3.340	3.3483	0.2987	0.00036	0.0912	3.2566	3.4391
90	1444	3.301	3.3558	0.2980	0.00037	0.0915	3.2642	3.4472
91	1417	3.329	3.3794	0.2959	0.00040	0.0922	3.2873	3.4717
92	1396	3.375	3.3988	0.2942	0.00044	0.0928	3.3062	3.4919
93	1367	3.424	3.4273	0.2918	0.00053	0.0938	3.3332	3.5208
94	1501	3.291	3.3110	0.3020	0.00040	0.0903	3.2209	3.4016
95	1406	3.399	3.3894	0.2950	0.00042	0.0925	3.2973	3.4824
96	1394	3.428	3.4007	0.2941	0.00045	0.0929	3.3073	3.4931
97	1410	3.369	3.3858	0.2954	0.00041	0.0924	3.2929	3.4776
98	1391	3.395	3.4036	0.2938	0.00045	0.0930	3.3107	3.4967
99	1445	3.332	3.3550	0.2981	0.00037	0.0914	3.2631	3.4460
100	1490	3.383	3.3192	0.3013	0.00038	0.0905	3.2285	3.4094
101	1460	3.325	3.3426	0.2992	0.00036	0.0911	3.2512	3.4333
102	1424	3.352	3.3731	0.2965	0.00039	0.0920	3.2807	3.4647
103	1424	3.355	3.3731	0.2965	0.00039	0.0920	3.2807	3.4647
104	1452	3.346	3.3491	0.2986	0.00036	0.0913	3.2577	3.4402
105	1425	3.355	3.3723	0.2965	0.00039	0.0920	3.2807	3.4647
106	1446	3.336	3.3541	0.2981	0.00037	0.0914	3.2631	3.4460
107	1411	3.398	3.3848	0.2954	0.00041	0.0924	3.2929	3.4776
108	1455	3.321	3.3467	0.2988	0.00036	0.0912	3.2555	3.4379
109	1442	3.344	3.3575	0.2978	0.00037	0.0915	3.2664	3.4495
110	1415	3.379	3.3812	0.2958	0.00041	0.0922	3.2884	3.4729
111	1398	3.416	3.3969	0.2944	0.00044	0.0928	3.3040	3.4895
112	1415	3.380	3.3812	0.2958	0.00041	0.0922	3.2884	3.4729
113	1417	3.376	3.3794	0.2959	0.00040	0.0922	3.2873	3.4717
114	1396	3.399	3.3988	0.2942	0.00044	0.0928	3.3062	3.4919
115	1378	3.420	3.4163	0.2927	0.00049	0.0934	3.3230	3.5099
116	1357	3.460	3.4375	0.2909	0.00057	0.0943	3.3433	3.5319
117	1410	3.370	3.3858	0.2954	0.00041	0.0924	3.2929	3.4776
118	1396	3.407	3.3988	0.2942	0.00044	0.0928	3.3062	3.4919
119	1378	3.441	3.4163	0.2927	0.00049	0.0934	3.3230	3.5099
120	1410	3.399	3.3858	0.2954	0.00041	0.0924	3.2929	3.4776
121	1418	3.395	3.3785	0.2960	0.00040	0.0922	3.2862	3.4705
122	1411	3.408	3.3848	0.2954	0.00041	0.0924	3.2929	3.4776
123	1408	3.404	3.3876	0.2952	0.00042	0.0924	3.2951	3.4800
124	1408	3.404	3.3876	0.2952	0.00042	0.0924	3.2951	3.4800
125	1408	3.404	3.3876	0.2952	0.00042	0.0924	3.2951	3.4800
126	1408	3.404	3.3876	0.2952	0.00042	0.0924	3.2951	3.4800

Obs	Thick	Resist	Pred	Xbeta	Std	95%PI Half-Width	L95% Predict	U95% Predict
127	1407	3.370	3.3885	0.2951	0.00042	0.0925	3.2962	3.4812
128	1479	3.285	3.3275	0.3005	0.00036	0.0907	3.2371	3.4185
129	1395	3.407	3.3998	0.2941	0.00045	0.0929	3.3073	3.4931
130	1354	3.350	3.4407	0.2906	0.00058	0.0944	3.3468	3.5356
131	1394	3.407	3.4007	0.2941	0.00045	0.0929	3.3073	3.4931
132	1495	3.263	3.3154	0.3016	0.00039	0.0904	3.2252	3.4061
133	1434	3.335	3.3644	0.2972	0.00038	0.0917	3.2730	3.4565
134	1523	3.278	3.2954	0.3034	0.00049	0.0901	3.2059	3.3861
135	1510	3.243	3.3045	0.3026	0.00043	0.0902	3.2145	3.3949
136	1450	3.305	3.3508	0.2984	0.00036	0.0913	3.2599	3.4425
137	1452	3.319	3.3491	0.2986	0.00036	0.0913	3.2577	3.4402
138	1556	3.191	3.2738	0.3055	0.00072	0.0902	3.1832	3.3635
139	1487	3.265	3.3214	0.3011	0.00037	0.0905	3.2306	3.4117
140	1474	3.372	3.3314	0.3002	0.00036	0.0908	3.2403	3.4219
141	1496	3.260	3.3147	0.3017	0.00039	0.0904	3.2242	3.4049
142	1550	3.232	3.2776	0.3051	0.00067	0.0901	3.1875	3.3677
143	1532	3.292	3.2893	0.3040	0.00054	0.0900	3.1994	3.3795
144	1456	3.330	3.3458	0.2989	0.00036	0.0912	3.2544	3.4368
145	1466	3.330	3.3378	0.2996	0.00036	0.0910	3.2468	3.4287
146	1512	3.300	3.3031	0.3027	0.00044	0.0902	3.2134	3.3938
147	1458	3.330	3.3442	0.2990	0.00036	0.0911	3.2533	3.4356
148	1463	3.311	3.3402	0.2994	0.00036	0.0910	3.2490	3.4310
149	1449	3.317	3.3516	0.2984	0.00036	0.0913	3.2599	3.4425
150	1536	3.231	3.2867	0.3043	0.00057	0.0900	3.1962	3.3763
151	1526	3.220	3.2934	0.3036	0.00051	0.0901	3.2037	3.3839
152	1455	3.319	3.3467	0.2988	0.00036	0.0912	3.2555	3.4379
153	1443	3.310	3.3567	0.2979	0.00037	0.0915	3.2653	3.4483
154	1452	3.320	3.3491	0.2986	0.00036	0.0913	3.2577	3.4402
155	1506	3.261	3.3074	0.3024	0.00042	0.0902	3.2167	3.3971
156	1460	3.255	3.3426	0.2992	0.00036	0.0911	3.2512	3.4333
157	1450	3.322	3.3508	0.2984	0.00036	0.0913	3.2599	3.4425
158	1540	3.249	3.2840	0.3045	0.00060	0.0901	3.1940	3.3741
159	1460	3.303	3.3426	0.2992	0.00036	0.0911	3.2512	3.4333
160	1497	3.250	3.3140	0.3018	0.00039	0.0903	3.2231	3.4038
161	1536	3.240	3.2867	0.3043	0.00057	0.0900	3.1962	3.3763
162	1468	3.303	3.3362	0.2997	0.00036	0.0909	3.2457	3.4276

CHAPTER 5

MONITORING UNIFORMITY ACROSS PROCESS ZONES AND CHANGING PRODUCT SPECIFICATIONS USING REGRESSION ADJUSTED VARIABLES

Introduction

In some manufacturing processes, large capacity equipment may have several zones, with slight differences between zones adding an additional source of variability. An important objective is ensuring that uniformity between zones is maintained to the highest possible extent. Furthermore, due to product delivery schedules, it may not be feasible to dedicate a single machine to a particular product formulation. Process attributes may need to be adjusted from run to run to meet target product specifications.

Considering each zone as a variable, scenarios where it is likely that a loss of uniformity occurs in a single zone at a time fit the key assumption made by Hawkins (1991) in proposing process monitoring with regression adjusted variables. While this method was shown to have degraded performance when simultaneous shifts of similar size in correlated variables occured, this situation does not represent a loss of uniformity. With uniformity as a principle interest, it is desirable to detect model-void shifts quickly.

Due to possible changes in product specification between runs, the regression adjustment must involve additional terms indicating which product specification is associated with the zone thickness measurements. The interest in this chapter is in forming a single model for each zone that assesses uniformity with respect to the other zones, and provides information on possible uniformity dispersion effects with respect to the product specification.

Example

A particular semi-conductor diffusion furnace has three zones defined by proximity to separate heating elements within the furnace. A single large heating element is insufficient to provide uniform heating throughout the furnace interior. Layer thickness in each zone is an important measure observed after furnace run completion. There are several possible mechanisms that may cause the thickness observed in one zone to move away from that observed in the other zones. A degraded thermocouple, controller channel, or heating element may result in a non-uniform temperature between one zone and the others. Furthermore, depending on the product loaded the desired thickness may be 400, 1400, 1500, or 2500 times a constant measurement unit. For convenience, product formulation specifications will be referred to as "recipes."

To represent four recipes, three indicator variables are required as regressors in each model in addition to the two continuous variables representing the other two zones. The models are:

$$\begin{aligned} x_1 &= \beta_1 d_1 + \beta_2 d_2 + \beta_3 d_3 + \beta_4 x_2 + \beta_5 x_3 + \varepsilon \\ x_2 &= \beta_6 d_1 + \beta_7 d_2 + \beta_8 d_3 + \beta_9 x_1 + \beta_{10} x_3 + \varepsilon \\ x_3 &= \beta_{11} d_1 + \beta_{12} d_2 + \beta_{13} d_3 + \beta_{14} x_1 + \beta_{15} x_2 + \varepsilon \end{aligned} \tag{5-1}$$

where x_i = Thickness in Zone i and the indicator variables, d_j, are coded such that:

Recipe	d_1	d_2	d_3
400	0	0	0
1400	1	0	0
1500	0	1	0
2500	0	0	1

A total of 894 observations of each zone thickness are available: 186 from the 400 recipe, 516 from the 1400 recipe, 57 from the 1500 recipe, and 135 from the 2500 recipe (the raw data is included in Appendix 5.A). Observations identified as nonconforming by process engineers were eliminated from the originally provided data set -- the intent being to assess variability in uniformity even when process is thought to be in a good state.

The fitted equations (using least squares estimates) associated with the models in (5-1) are:

$$
\begin{aligned}
\hat{x}_1 &= 1.82d_1 + 17.3d_2 + 44.9d_3 + 0.32x_2 + 0.67x_3 \\
\hat{x}_2 &= 16.8d_1 + 9.0d_2 + 20.0d_3 + 0.15x_1 + 0.85x_3 \\
\hat{x}_3 &= -7.37d_1 - 6.27d_2 - 21.1d_3 + 0.27x_1 + 0.74x_2
\end{aligned}
\tag{5-2}
$$

Figure 5-1 contains a boxplot of the zone 1 residuals ($x_1 - \hat{x}_1$) grouped by recipe, which shows an interesting process phenomenon -- since these residuals represent zone one departure from its usual relationship to zones two and three, we see a recipe dispersion effect in uniformity, especially for the 1400 recipe. In this process it may be reasonable to expect larger variability within a zone as the target thickness increases, and hence perhaps a tendency to observe this across zones as well; but the dispersion effect for the 1400 recipe seems to violate the more mild tendency for this to happen when the progression from 400 to 1500 to 2500 is observed. Possible causes for the excess uniformity dispersion in the 1400 recipe level should be investigated.

Figure 2 contains confidence intervals around the residual variance estimates as well as a Bartlett's test for Homogeneity of variance, indicating the dispersion effect noticed in Figure 1 is statistically significant.

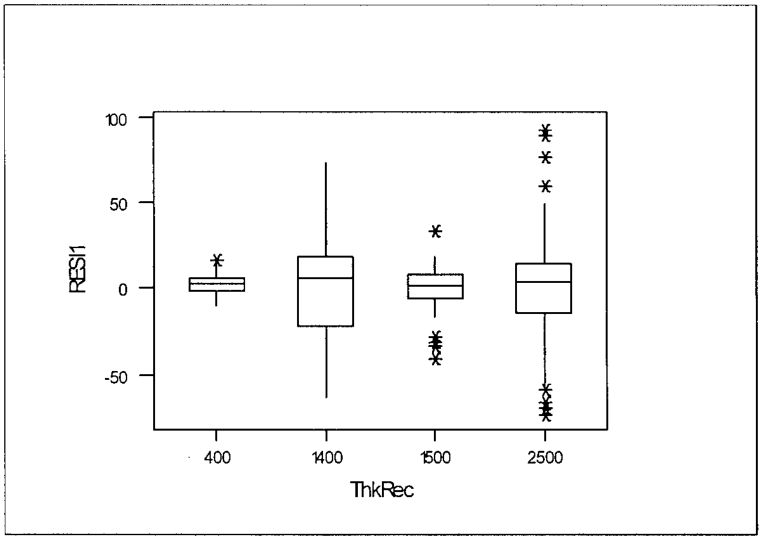


Figure 5-1. Box and whisker plots of zone one residuals by recipe.

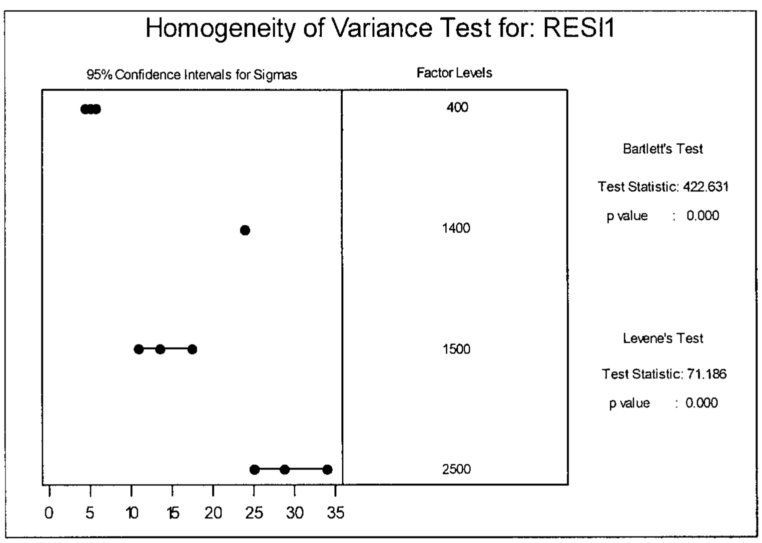


Figure 5-2. Homogeneity of variance test for zone 1 residuals.

Figures 5-3 and 5-4 show similar dispersion effects for zones two and three, respectively.

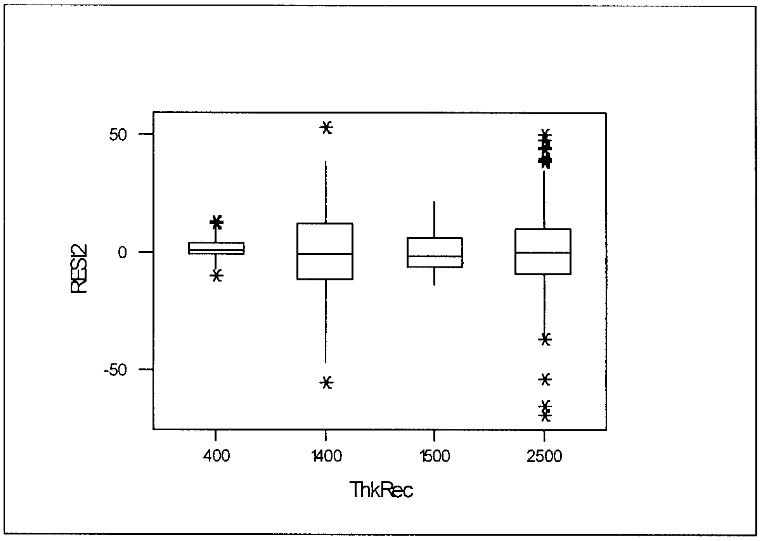


Figure 5-3. Box and whisker plots of zone two residuals by recipe.

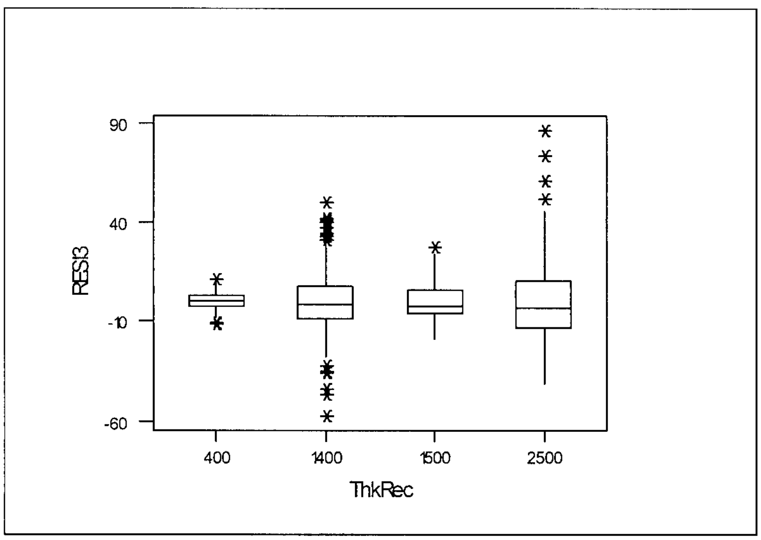


Figure 5-4. Box and whisker plots of zone three residuals by recipe.

Table 5-1 presents seven points from the process and clearly shows how the regression adjusted variables will operate. Observation number 18 shows a point that is very close to target in all three zones -- the standardized residuals remain small as they should.

Table 5-1. Summary statistics on selected process observations.

Obs	ThkRec	Z1Thk	Z2Thk	Z3Thk	Range	Z1Std	Z2Std	Z3Std	SRES1	SRES2	SRES3
18	400	395.47	397.13	399.32	3.85	-0.63	-0.50	-0.38	0.02	0.05	0.10
483	1400	1391.60	1418.70	1350.50	68.2	0.22	0.94	-2.23	1.37	3.60	-4.19
845	2500	2537.30	2429.60	2422.10	115.20	1.07	-1.94	-1.79	4.13	-1.05	-1.66
858	2500	2528.50	2420.40	2409.20	119.30	0.79	-2.25	-2.20	4.26	-0.84	-1.94
878	2500	2429.60	2423.30	2497.90	74.60	-2.43	-2.15	0.60	-3.11	-4.77	6.29
882	400	390.51	389.35	393.56	4.21	-1.49	-1.72	-1.29	0.09	-0.10	0.20
889	1400	1439.50	1386.80	1371.60	67.90	2.05	-0.83	-1.16	3.38	-0.25	-1.91

Variable Definitions:

Obs = Observation Number

ThkRec = Thickness Recipe

Z1Thk = Zone 1 Thickness

Z2 Thk = Zone 2 Thickness

Z3Thk = Zone 3 Thickness

Range = Thickness Range Across All Zones

Z1Std = Standardized Zone 1 Thickness

Z2Std = Standardized Zone 2 Thickness

Z3Std = Standardized Zone 3 Thickness

SRES1 = Standardized Zone 1 Residual

SRES2 = Standardized Zone 2 Residual

SRES3 = Standardized Zone 3 Residual

Observation 882 shows a situation where observed thicknesses in all three zones are smaller than usual but remain similar with respect to their usual relationship -- the standardized residuals remain small since uniformity is maintained (though it's probably important to pick up a shift in the overall mean, a topic that will be discussed later).

Observation 845 shows a situation where the zone one observation is substantially higher than the other zones -- the standardized residual in zone 1 is very large (4.13) indicating the problem quite clearly.

Observation 483 has zone 1 remaining near target, zone two moderately larger than target, and zone 3 substantially smaller than target -- this situation is also accurately shown in the standardized residuals.

A very interesting comparison is provided between observations 858 and 878. Both observations have two zones that are substantially lower than target (different two in each observation), with the remaining zone slightly above target -- magnitudes are similar, only the zones involved are changed. Even so, the zone 3 standardized residual for observation 878 is much larger than the zone 1 residual for observation 858 -- this is indicative of the historical relationship information between zones coming into play. Another important point lies in this comparison. The range across zones for observation 858 is 1.6 times the range across zones for observation 878, yet observation 878 contains the largest standardized residual in any of the zones. This indicates regression-adjusted monitoring has the potential to be more sensitive than a range-based control method for this application.

Looking back over these examples, the regression adjusted variables detection ability is not just limited to single zone shifts, but depends on the violation of normal relationships between the zones -- a mechanism very well suited to monitoring uniformity.

In all cases shown, the standardized thickness indicate that all observations would be well inside control limits on univariate individuals control charts -- separately charting the zone thickness would provide very little information on uniformity except under very large shifts.

Other Considerations

Observation 882 demonstrated that the residuals remain small when the overall mean thickness changes uniformly across zones. This situation is also important to detect. Using a T^2, MCUSUM, or MEWMA "in parallel" with monitoring the regression adjusted variables would signal the uniform changes in the mean. To keep the number of charts being monitored to a minimum, a "ZNO" group chart presented by Hawkins (1991) (discussed in Chapter 2) could be used along with, for example, an MCUSUM.

One disadvantage of monitoring with residuals from the model in (5-1) is that the residuals for zone one are not indepent of those for zones two and three. Simulation must be used to determine appropriate control chart parameters as discussed by Hawkins (1991).

Least-squares regression requires assumptions of constance variance. Once the dispersion problem associated with recipe 1400 was corrected, an increasing trend in dispersion would still be present based on observing the trend that existed between

recipes 400, 1500, and 2500. A variance stabilizing transformation should be used to correct the problem prior to fitting the models that will be used.

Alternatively, one may standardize the observations, using the mean and standard deviation observed for each recipe within each zone. The models become:

$$\begin{aligned} z_1 &= \beta_1 z_2 + \beta_2 z_3 + \varepsilon \\ z_2 &= \beta_3 z_1 + \beta_4 z_3 + \varepsilon \\ z_3 &= \beta_5 z_2 + \beta_6 z_2 + \varepsilon \end{aligned} \tag{5-3}$$

where, $z_{ijk} = \frac{x_{ijk} - \overline{x}_{ij}}{s_{ij}}$ for the i^{th} zone, j^{th} recipe and k^{th} observation. Figure 5-5 contains a boxplot of the zone one residuals formed in this manner grouped by recipe. This figure demonstrates that standardizing in the manner of (5-3) largely masks the dispersion effects. The residual variance confidence intervals and Bartlett's test in Figure 5-6 no longer indicate non-homogenous variance. While this may be a reasonable approach for continued monitoring once the dispersion has been reduced to an acceptable level, we would have missed the important information regarding the uniformity dispersion problem in recipe 1400 had we done this at the beginning.

Summary and Conclusions

Maintaining uniformity across equipment zones is an important process monitoring objective. Hawkins (1991) regresssion adjustment method was shown to be more sensitive to shifts that violate normal relationships between variables than were traditional methods applied to original observations. By treating zones as variables,

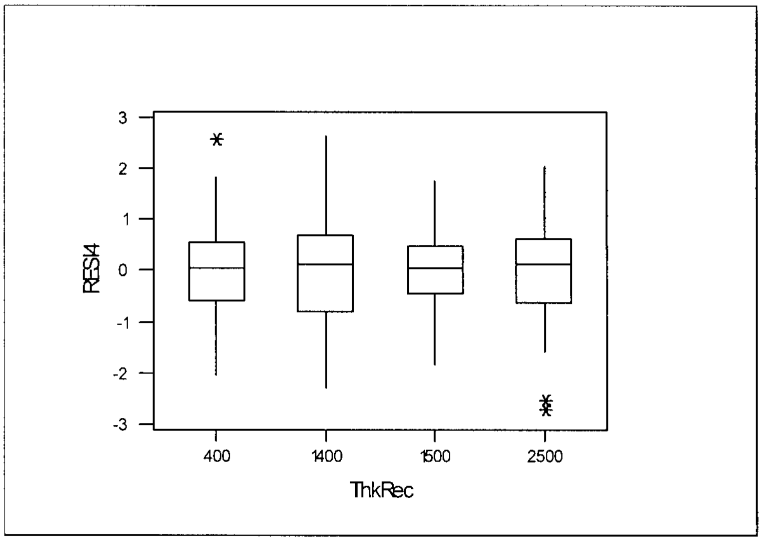


Figure 5-5. Box and whisker plot of zone one residuals from model using "recipe-standardized" variables.

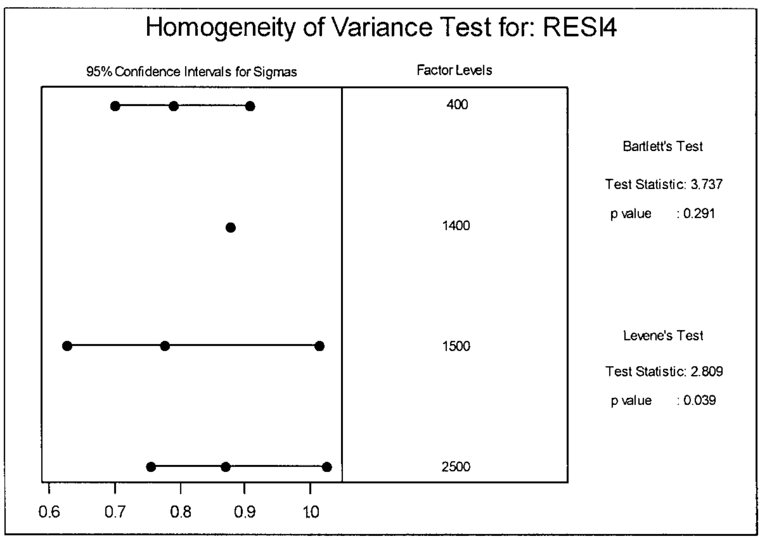


Figure 5-6. Homogeneity of variance test for zone one residuals from model using "recipe-standardized" variables.

monitoring zone uniformity fits well into Hawkins (1991) framework. Additional factors such as product specification "recipe" may be incorporated as additional covariates. Doing so clearly identified dispersion effects across recipe formulations that are masked when variables are standardized instead as a means for monitoring differing targets on the same machine. A significant disadvantage is that zone residuals are not independent of

each other, so simulation is required to determine appropriate control chart parameters. "Parallel" monitoring with a directionally invariant procedure such as a T^2, MCUSUM or MEWMA on raw measures is important for detecting an overall mean shift that occurs uniformly across all zones.

APPENDIX 5.A

Zone Thickness Data

Obs	Rec	Z1Thk	Z2Thk	Z3Thk	Obs	Rec	Z1Thk	Z2Thk	Z3Thk
1	1500	1462.70	1445.60	1444.30	46	1400	1430.50	1408.30	1405.30
2	1400	1398.30	1384.10	1375.20	47	1400	1429.30	1408.00	1404.30
3	1400	1399.00	1380.50	1369.80	48	1400	1411.80	1405.80	1408.30
4	1400	1403.00	1379.40	1367.30	49	1400	1406.80	1396.30	1401.10
5	1400	1406.50	1379.10	1369.50	50	1400	1424.60	1396.60	1396.80
6	1400	1421.10	1391.20	1396.10	51	1500	1499.20	1482.10	1482.50
7	1400	1425.10	1393.10	1400.60	52	1500	1504.90	1487.20	1487.80
8	1400	1418.80	1409.90	1414.70	53	1400	1408.70	1406.90	1407.80
9	1500	1506.20	1488.40	1487.90	54	1400	1415.60	1391.80	1404.80
10	1400	1413.30	1403.20	1409.00	55	1400	1421.20	1397.00	1418.90
11	1400	1422.20	1402.20	1400.30	56	1400	1406.70	1392.80	1421.70
12	1400	1407.80	1396.10	1392.00	57	1400	1400.50	1402.30	1414.40
13	1400	1419.40	1408.50	1405.70	58	1400	1401.10	1394.80	1412.70
14	1400	1406.90	1405.20	1407.90	59	1400	1398.40	1401.40	1410.10
15	1400	1412.30	1412.50	1413.30	60	1400	1401.50	1391.90	1415.90
16	1400	1411.40	1412.60	1415.70	61	1400	1396.50	1384.50	1407.60
17	1400	1434.80	1420.80	1425.50	62	1400	1402.60	1383.80	1400.00
18	400	395.47	397.13	399.32	63	1400	1390.20	1410.20	1390.00
19	400	395.66	396.20	398.94	64	1400	1401.80	1389.30	1405.80
20	400	397.83	396.35	399.62	65	400	387.04	388.75	392.36
21	400	400.39	399.35	400.46	66	1500	1523.00	1508.60	1510.30
22	1400	1443.00	1420.30	1423.10	67	1400	1394.20	1395.20	1402.70
23	1500	1512.90	1492.60	1493.60	68	1400	1406.40	1398.80	1418.50
24	1400	1434.60	1409.80	1419.90	69	1400	1405.70	1408.90	1416.00
25	400	397.54	394.25	396.43	70	1400	1411.60	1390.30	1413.00
26	1400	1446.00	1421.60	1423.00	71	2500	2469.60	2467.20	2485.50
27	1400	1424.30	1401.10	1404.40	72	1400	1377.50	1381.10	1388.00
28	1400	1425.30	1403.40	1405.00	73	1400	1403.10	1376.40	1384.10
29	1500	1492.60	1476.20	1475.60	74	1400	1397.30	1388.10	1406.40
30	1500	1497.90	1477.20	1474.80	75	1400	1392.70	1373.40	1410.50
31	1500	1497.40	1476.40	1474.10	76	1400	1394.00	1373.60	1400.20
32	1400	1422.40	1397.70	1399.90	77	400	384.02	384.30	388.47
33	2500	2497.00	2482.00	2484.00	78	1400	1390.90	1377.40	1402.20
34	1400	1427.50	1400.20	1398.70	79	1400	1384.90	1386.40	1394.70
35	1400	1417.90	1401.00	1398.00	80	1400	1378.40	1382.60	1395.50
36	1400	1416.20	1395.30	1395.60	81	1400	1372.80	1380.40	1397.40
37	400	394.93	390.58	393.19	82	1400	1406.00	1385.90	1426.00
38	400	398.45	398.01	399.79	83	400	402.51	404.07	414.08
39	400	397.25	395.29	397.60	84	1400	1400.50	1402.90	1412.60
40	400	397.31	395.57	397.65	85	1400	1414.40	1396.00	1415.30
41	400	393.88	392.79	395.35	86	1400	1405.10	1391.30	1408.60
42	400	392.25	392.07	395.30	87	1400	1401.40	1392.80	1411.00
43	400	392.51	392.65	395.72	88	1400	1402.30	1393.50	1413.50
44	1400	1404.10	1407.90	1431.70	89	1400	1407.90	1393.90	1411.70
45	1400	1422.00	1405.30	1403.70	90	1400	1401.40	1387.00	1415.60

Obs	Rec	Z1Thk	Z2Thk	Z3Thk	Obs	Rec	Z1Thk	Z2Thk	Z3Thk
91	400	409.34	406.14	408.76	136	2500	2520.70	2490.10	2485.50
92	1500	1489.70	1473.50	1482.40	137	1400	1397.10	1389.30	1386.70
93	1400	1404.60	1385.70	1402.80	138	2500	2536.10	2521.20	2534.20
94	1400	1402.40	1385.70	1401.80	139	400	403.67	402.82	402.88
95	1400	1404.60	1393.60	1410.50	140	2500	2481.10	2462.10	2458.90
96	1400	1386.80	1371.90	1395.00	141	1400	1403.40	1384.30	1388.60
97	400	402.39	400.90	403.82	142	1400	1388.50	1383.60	1392.90
98	1400	1391.40	1391.60	1394.20	143	400	406.26	406.28	407.32
99	1400	1391.20	1375.00	1394.30	144	1400	1396.50	1395.10	1396.40
100	2500	2451.20	2434.90	2447.00	145	1400	1402.10	1395.50	1397.80
101	1400	1388.30	1391.10	1373.00	146	1400	1389.80	1389.40	1401.90
102	1400	1410.00	1389.00	1402.00	147	1400	1364.90	1364.00	1379.20
103	1400	1398.30	1383.30	1399.10	148	1500	1481.30	1473.40	1481.70
104	1400	1404.20	1388.00	1410.00	149	1400	1401.70	1394.60	1396.50
105	1400	1399.60	1374.40	1408.80	150	1400	1401.90	1394.70	1395.40
106	400	403.30	405.58	409.31	151	1400	1400.70	1397.40	1400.90
107	1400	1409.80	1397.30	1427.40	152	1400	1393.40	1385.10	1387.60
108	1400	1394.20	1378.20	1420.80	153	400	401.53	400.91	401.80
109	1400	1392.70	1388.50	1395.80	154	1400	1397.40	1382.70	1383.60
110	1400	1390.90	1384.20	1390.40	155	2500	2460.40	2444.80	2443.60
111	1400	1398.00	1388.50	1402.50	156	400	393.24	393.45	400.93
112	1400	1403.30	1394.70	1400.60	157	1400	1395.70	1380.80	1382.40
113	1500	1476.50	1472.70	1488.20	158	1400	1390.80	1373.90	1375.80
114	1400	1390.60	1386.60	1416.10	159	1400	1389.30	1383.50	1392.20
115	1400	1390.00	1391.00	1405.40	160	400	391.51	396.73	405.70
116	1400	1397.90	1391.70	1402.60	161	1400	1389.60	1378.60	1393.10
117	2500	2517.40	2497.40	2514.60	162	1400	1391.70	1385.00	1394.10
118	400	396.00	397.00	404.94	163	1400	1395.00	1386.50	1398.80
119	1400	1399.60	1385.00	1392.80	164	1400	1391.60	1384.00	1391.50
120	1400	1382.10	1398.70	1407.00	165	1400	1397.00	1396.70	1395.60
121	1400	1393.90	1392.40	1406.50	166	1400	1398.40	1396.70	1398.60
122	2500	2482.10	2465.50	2471.30	167	2500	2484.80	2462.20	2457.10
123	1400	1393.90	1386.50	1394.80	168	1400	1401.40	1392.60	1389.50
124	1400	1394.30	1389.90	1397.40	169	1400	1403.20	1395.30	1393.10
125	2500	2484.60	2464.80	2468.70	170	1400	1385.70	1381.70	1390.00
126	1400	1397.20	1393.30	1391.20	171	400	402.22	401.05	402.81
127	1400	1394.10	1388.00	1394.10	172	1400	1399.00	1385.80	1383.10
128	1400	1403.50	1395.90	1396.40	173	400	404.90	402.65	403.88
129	400	405.81	403.56	406.33	174	2500	2475.30	2457.70	2451.50
130	1400	1401.00	1389.30	1391.70	175	1400	1397.90	1386.80	1384.60
131	1400	1398.30	1390.60	1394.40	176	1500	1483.80	1473.50	1487.70
132	1500	1501.30	1487.50	1495.30	177	1400	1384.20	1392.90	1393.00
133	400	406.99	405.72	408.07	178	2500	2505.40	2482.20	2497.20
134	400	408.59	407.20	410.03	179	2500	2479.50	2463.50	2465.40
135	2500	2497.10	2473.80	2468.60	180	1400	1389.60	1386.40	1391.00

Obs	Rec	Z1Thk	Z2Thk	Z3Thk	Obs	Rec	Z1Thk	Z2Thk	Z3Thk
181	1400	1381.10	1384.40	1391.80	226	1400	1399.90	1412.00	1408.50
182	1400	1387.20	1388.90	1395.80	227	1400	1393.20	1406.60	1396.70
183	1400	1386.60	1389.20	1393.40	228	1500	1501.50	1496.60	1503.50
184	1400	1387.70	1385.70	1388.80	229	1400	1397.60	1413.80	1400.60
185	1400	1384.00	1378.90	1384.50	230	1400	1390.50	1415.60	1415.80
186	400	402.38	401.72	403.08	231	1400	1387.70	1418.00	1417.80
187	1400	1376.80	1375.40	1385.90	232	400	405.01	397.07	406.84
188	1400	1386.30	1382.30	1390.60	233	400	404.13	402.93	409.59
189	1400	1386.70	1378.40	1389.40	234	1400	1414.50	1408.00	1395.90
190	1400	1373.90	1371.40	1375.10	235	400	402.91	400.93	410.04
191	1400	1371.90	1368.50	1375.30	236	400	410.13	403.84	409.26
192	400	403.74	402.30	405.61	237	1400	1406.90	1426.50	1434.70
193	1400	1380.30	1372.10	1375.40	238	1400	1408.80	1419.40	1417.70
194	1400	1382.10	1373.30	1378.10	239	1400	1405.40	1424.90	1422.80
195	1400	1378.70	1368.10	1372.30	240	1400	1404.40	1425.20	1430.60
196	1400	1374.10	1369.80	1372.40	241	1500	1489.00	1476.70	1473.80
197	400	399.97	398.90	402.11	242	1400	1399.90	1417.30	1414.80
198	1400	1375.50	1371.40	1375.00	243	2500	2520.20	2499.20	2487.50
199	2500	2447.90	2433.20	2427.80	244	1400	1397.50	1411.50	1411.10
200	1400	1377.40	1368.20	1373.00	245	1500	1489.10	1483.70	1479.60
201	2500	2452.60	2432.30	2425.30	246	1400	1401.60	1410.50	1410.20
202	1400	1384.50	1376.90	1383.10	247	1400	1376.50	1421.60	1412.60
203	1500	1462.50	1453.10	1457.80	248	1400	1371.30	1420.00	1405.50
204	1400	1382.60	1373.00	1370.80	249	1500	1494.50	1491.50	1492.70
205	1400	1383.90	1374.80	1372.30	250	400	404.53	399.26	407.16
206	400	408.29	406.44	406.56	251	2500	2493.80	2472.70	2465.50
207	1400	1407.80	1400.70	1401.10	252	1400	1372.90	1414.80	1404.30
208	1400	1414.00	1409.30	1405.20	253	1400	1377.40	1416.50	1406.40
209	1400	1408.30	1399.10	1394.20	254	1400	1384.10	1424.10	1426.80
210	2500	2494.60	2477.30	2467.80	255	1400	1434.40	1408.00	1412.50
211	1400	1403.40	1398.40	1396.10	256	1400	1375.30	1419.40	1409.00
212	400	403.46	401.50	402.06	257	400	409.26	397.59	404.40
213	1400	1399.10	1385.10	1388.70	258	1400	1378.20	1420.70	1413.20
214	2500	2514.60	2490.00	2480.20	259	1400	1370.00	1417.40	1403.00
215	2500	2500.70	2476.10	2464.80	260	1400	1377.50	1414.00	1392.50
216	400	401.40	399.00	399.09	261	1400	1383.90	1438.60	1425.60
217	1400	1388.10	1376.80	1370.70	262	1400	1384.90	1434.10	1428.90
218	1400	1384.80	1373.60	1367.60	263	1400	1382.00	1435.20	1426.90
219	400	400.29	397.01	398.26	264	400	412.13	404.00	409.00
220	2500	2498.50	2478.30	2472.50	265	1400	1377.80	1435.80	1427.60
221	1400	1377.30	1370.10	1368.70	266	2500	2535.80	2512.40	2506.80
222	1400	1437.50	1401.60	1395.70	267	1400	1432.90	1423.10	1416.10
223	400	401.81	395.57	401.50	268	400	403.03	400.28	401.13
224	1400	1388.20	1415.10	1403.30	269	1400	1407.70	1403.40	1403.10
225	1400	1398.00	1412.50	1401.80	270	400	404.11	401.63	403.92

Obs	Rec	Z1Thk	Z2Thk	Z3Thk	Obs	Rec	Z1Thk	Z2Thk	Z3Thk
271	400	405.79	401.16	407.05	316	1400	1367.40	1414.30	1393.30
272	2500	2531.80	2511.70	2506.20	317	1400	1405.00	1444.00	1422.00
273	400	405.02	402.19	406.68	318	1400	1392.00	1437.00	1407.00
274	1400	1438.90	1427.80	1420.80	319	1400	1411.90	1408.00	1407.90
275	1400	1367.00	1402.90	1391.20	320	1400	1408.20	1404.00	1404.20
276	400	401.90	399.17	400.54	321	2500	2523.30	2486.40	2479.50
277	1400	1386.20	1429.10	1419.70	322	1400	1361.70	1417.10	1409.60
278	1400	1390.80	1444.90	1445.80	323	1400	1363.80	1420.20	1399.90
279	1400	1387.00	1429.10	1410.60	324	1400	1365.80	1425.30	1397.60
280	1400	1446.30	1432.10	1431.60	325	400	397.74	403.30	407.25
281	1400	1436.30	1424.20	1420.20	326	1400	1355.30	1419.20	1394.60
282	400	408.58	406.86	407.60	327	2500	2531.90	2488.90	2481.20
283	2500	2551.90	2527.90	2522.40	328	1400	1426.10	1416.30	1413.80
284	1400	1420.30	1413.80	1412.50	329	1500	1511.00	1518.50	1509.30
285	1500	1492.30	1503.80	1498.00	330	1400	1428.10	1421.00	1417.60
286	2500	2548.70	2526.00	2520.50	331	400	402.33	411.18	404.93
287	1500	1481.20	1469.80	1469.20	332	1400	1366.00	1420.50	1390.80
288	1400	1378.40	1431.90	1413.20	333	1400	1416.40	1419.80	1412.10
289	1400	1385.20	1430.50	1416.80	334	1400	1429.50	1421.20	1402.90
290	1400	1435.80	1427.20	1418.80	335	1400	1412.80	1414.10	1406.70
291	1500	1521.80	1513.30	1520.40	336	1500	1507.00	1511.60	1499.50
292	1400	1386.00	1429.60	1407.30	337	400	397.01	408.19	403.81
293	1400	1381.00	1435.50	1418.70	338	1400	1428.90	1421.80	1401.20
294	1500	1536.80	1533.40	1533.30	339	1500	1501.80	1511.50	1497.20
295	1400	1386.10	1439.30	1431.90	340	1500	1497.90	1497.50	1490.70
296	1400	1444.60	1424.60	1422.80	341	2500	2521.40	2507.80	2498.20
297	1400	1444.40	1428.90	1420.20	342	400	395.56	408.03	400.80
298	400	408.97	404.39	408.19	343	1400	1407.00	1411.30	1402.20
299	1400	1389.30	1431.80	1415.90	344	1400	1413.90	1423.70	1410.50
300	1400	1375.70	1418.30	1415.30	345	1400	1431.00	1425.80	1412.10
301	400	395.48	396.71	402.57	346	400	400.66	411.01	403.72
302	1400	1372.90	1420.20	1399.40	347	1400	1414.20	1418.90	1405.70
303	1400	1375.70	1431.20	1436.20	348	1400	1422.30	1422.80	1405.50
304	1400	1372.90	1417.10	1397.40	349	400	391.98	404.90	403.00
305	2500	2549.30	2525.20	2514.00	350	1400	1418.10	1421.10	1409.00
306	1400	1367.80	1413.90	1386.90	351	1400	1400.90	1407.10	1397.80
307	1400	1375.90	1413.60	1394.50	352	1500	1491.70	1498.70	1504.60
308	1400	1366.20	1413.40	1400.50	353	1400	1432.20	1426.20	1397.00
309	1400	1366.60	1413.00	1396.70	354	2500	2515.40	2481.40	2470.90
310	1400	1377.10	1422.20	1396.10	355	400	391.58	406.46	399.32
311	1400	1366.90	1413.00	1395.70	356	1400	1415.90	1422.60	1407.80
312	1400	1365.20	1411.80	1395.20	357	1400	1353.20	1416.40	1392.20
313	1400	1374.10	1435.30	1456.30	358	1500	1495.50	1498.50	1500.10
314	1500	1517.40	1502.70	1496.90	359	1400	1422.40	1421.60	1408.10
315	2500	2530.20	2488.00	2481.60	360	2500	2530.80	2514.40	2493.80

Obs	Rec	Z1Thk	Z2Thk	Z3Thk	Obs	Rec	Z1Thk	Z2Thk	Z3Thk
361	400	404.80	404.79	406.68	406	1400	1359.70	1415.70	1391.60
362	1400	1427.90	1418.70	1412.30	407	1400	1435.40	1428.60	1413.80
363	1400	1423.90	1412.70	1410.30	408	1400	1364.90	1418.90	1402.30
364	400	394.27	396.36	402.11	409	2500	2515.90	2501.60	2476.40
365	1400	1424.70	1413.00	1414.90	410	400	400.86	410.56	403.94
366	1500	1515.90	1502.40	1507.90	411	1400	1356.90	1418.10	1392.00
367	1400	1426.00	1416.90	1410.60	412	1400	1353.30	1418.80	1394.90
368	1400	1423.70	1416.40	1412.30	413	1400	1419.40	1414.30	1406.20
369	1400	1361.20	1418.90	1390.60	414	1400	1425.40	1419.50	1409.10
370	1500	1507.10	1513.30	1501.00	415	2500	2523.30	2499.70	2473.00
371	1500	1520.30	1523.90	1520.10	416	1400	1358.60	1413.90	1390.00
372	1400	1411.90	1408.90	1404.80	417	1400	1366.40	1428.40	1399.70
373	1400	1427.90	1420.40	1417.80	418	2500	2522.80	2488.30	2477.60
374	2500	2543.20	2524.50	2510.60	419	400	394.19	409.55	399.37
375	1400	1426.60	1414.30	1409.60	420	1400	1431.20	1439.50	1410.50
376	1400	1357.80	1412.90	1393.80	421	1400	1432.60	1434.40	1412.70
377	1400	1427.40	1420.40	1410.60	422	1400	1427.00	1428.60	1410.40
378	1400	1420.50	1410.00	1408.50	423	2500	2525.00	2507.10	2478.90
379	1400	1363.50	1422.00	1413.00	424	1400	1366.90	1425.70	1393.40
380	1500	1502.00	1501.00	1485.30	425	1400	1357.10	1421.80	1394.10
381	2500	2530.80	2516.40	2501.50	426	1400	1354.50	1426.20	1389.10
382	1400	1359.30	1417.80	1408.70	427	1400	1358.60	1418.50	1388.00
383	1500	1505.30	1511.90	1512.70	428	1400	1404.60	1408.50	1390.30
384	1400	1409.00	1411.40	1405.50	429	1400	1423.50	1415.80	1410.90
385	2500	2525.60	2511.30	2498.00	430	1400	1430.10	1420.40	1411.00
386	1400	1357.80	1415.40	1398.90	431	1400	1358.00	1418.60	1402.40
387	2500	2527.50	2511.00	2496.50	432	1400	1361.70	1412.80	1392.20
388	1400	1416.90	1418.60	1408.90	433	2500	2524.60	2499.60	2481.50
389	1400	1429.80	1421.30	1412.00	434	1400	1363.60	1413.50	1402.50
390	1400	1425.60	1417.50	1408.20	435	2500	2521.00	2492.50	2480.00
391	1400	1427.30	1426.90	1419.10	436	400	392.23	401.06	398.66
392	1400	1419.00	1416.10	1394.70	437	1500	1504.10	1504.20	1494.00
393	1400	1422.70	1418.00	1411.80	438	1400	1362.10	1418.00	1407.40
394	400	392.45	398.69	400.53	439	1500	1503.10	1501.60	1492.10
395	1400	1363.80	1415.70	1388.70	440	2500	2524.10	2486.60	2473.20
396	1400	1366.20	1420.00	1390.10	441	1400	1358.40	1416.20	1411.20
397	1400	1421.60	1419.00	1410.90	442	1400	1356.00	1421.50	1388.90
398	1400	1362.70	1415.20	1388.10	443	1400	1358.60	1418.00	1393.70
399	2500	2529.40	2508.60	2500.90	444	1500	1509.20	1508.50	1503.50
400	1400	1424.10	1414.20	1410.40	445	2500	2524.00	2501.10	2478.50
401	1400	1422.30	1424.50	1414.80	446	1400	1354.70	1411.70	1389.30
402	1400	1433.60	1427.10	1415.10	447	1500	1503.50	1500.40	1491.90
403	1400	1429.30	1428.50	1420.00	448	400	395.73	404.07	403.54
404	1400	1426.10	1426.00	1406.50	449	1400	1368.70	1428.00	1418.70
405	2500	2528.60	2504.70	2479.20	450	1400	1366.30	1424.10	1397.90

Obs	Rec	Z1Thk	Z2Thk	Z3Thk	Obs	Rec	Z1Thk	Z2Thk	Z3Thk
451	1400	1360.80	1419.40	1387.90	496	1400	1371.10	1423.70	1400.70
452	1400	1360.80	1412.90	1386.50	497	400	399.85	407.36	409.00
453	1400	1364.90	1412.30	1391.40	498	1400	1374.20	1425.30	1397.60
454	1400	1360.20	1413.10	1388.80	499	400	395.85	405.76	406.27
455	400	395.41	401.57	400.06	500	1400	1378.50	1429.70	1398.10
456	2500	2515.30	2498.40	2477.60	501	2500	2547.20	2528.30	2513.10
457	1400	1360.90	1418.90	1385.90	502	400	400.35	412.21	410.15
458	1400	1357.10	1418.90	1390.00	503	2500	2506.80	2530.70	2546.70
459	2500	2524.70	2493.00	2472.80	504	1400	1397.70	1460.50	1458.80
460	1400	1367.50	1425.10	1396.30	505	1400	1394.50	1452.80	1459.50
461	1400	1407.30	1405.40	1397.60	506	1400	1371.70	1432.50	1441.30
462	2500	2510.80	2491.90	2480.50	507	1400	1336.30	1399.00	1400.70
463	1400	1356.50	1419.20	1390.10	508	1400	1347.50	1413.10	1426.60
464	400	388.10	396.01	399.88	509	1400	1343.80	1401.00	1391.40
465	1400	1359.00	1412.10	1390.30	510	2500	2477.00	2475.50	2507.10
466	400	391.94	398.70	402.16	511	1400	1353.60	1412.60	1418.20
467	1400	1352.10	1418.30	1388.80	512	1500	1497.50	1495.30	1506.10
468	1400	1357.80	1416.60	1386.60	513	1400	1352.90	1415.40	1415.90
469	400	393.80	404.36	399.55	514	2500	2474.80	2468.50	2476.10
470	1400	1357.10	1415.40	1388.70	515	1400	1324.10	1390.70	1370.30
471	400	391.93	404.02	401.19	516	1400	1337.90	1393.00	1384.40
472	1400	1359.60	1418.80	1386.90	517	1400	1347.00	1412.60	1409.60
473	2500	2510.90	2489.40	2467.90	518	2500	2477.60	2442.50	2458.10
474	1400	1354.90	1414.50	1376.80	519	1500	1500.60	1488.10	1495.30
475	1400	1358.70	1414.30	1383.80	520	1400	1346.30	1391.10	1391.00
476	1400	1359.30	1410.60	1398.20	521	1400	1354.90	1398.90	1403.10
477	1400	1398.80	1391.70	1377.50	522	1500	1514.10	1521.80	1528.80
478	1400	1354.00	1414.20	1392.80	523	1400	1330.90	1379.00	1361.60
479	2500	2505.10	2495.90	2481.20	524	400	402.45	403.28	404.22
480	1400	1407.90	1411.50	1388.80	525	1400	1393.50	1396.20	1345.60
481	1400	1411.60	1401.70	1388.50	526	1500	1489.00	1480.80	1480.50
482	1400	1403.80	1403.80	1391.50	527	1400	1332.10	1374.60	1360.20
483	1400	1391.60	1418.70	1350.50	528	2500	2460.60	2447.20	2439.80
484	1500	1480.00	1488.60	1479.30	529	1400	1381.30	1371.30	1352.50
485	1400	1359.90	1418.80	1384.90	530	1400	1332.10	1392.90	1394.80
486	1400	1409.90	1401.00	1386.70	531	2500	2520.60	2505.80	2488.20
487	1400	1421.80	1406.30	1397.10	532	1400	1422.90	1427.90	1419.90
488	400	408.91	413.08	411.00	533	400	394.65	401.49	398.87
489	1400	1363.00	1417.30	1403.70	534	2500	2460.20	2449.20	2446.20
490	1400	1356.70	1413.50	1385.10	535	400	396.91	405.37	403.35
491	2500	2505.20	2484.00	2468.40	536	1400	1382.60	1372.80	1364.60
492	1400	1364.70	1428.70	1406.20	537	1400	1404.90	1396.20	1396.00
493	1400	1376.60	1430.50	1399.90	538	400	396.03	410.50	408.31
494	400	396.44	402.76	405.40	539	1500	1491.60	1482.00	1477.80
495	400	396.88	401.58	404.20	540	2500	2461.30	2450.40	2445.70

Obs	Rec	Z1Thk	Z2Thk	Z3Thk	Obs	Rec	Z1Thk	Z2Thk	Z3Thk
541	1400	1357.80	1401.30	1390.70	586	1400	1345.00	1398.00	1377.50
542	400	390.50	395.89	395.80	587	2500	2559.50	2539.60	2533.60
543	1400	1341.40	1399.60	1411.00	588	1400	1342.70	1398.10	1380.50
544	2500	2532.40	2503.90	2485.60	589	1400	1352.90	1414.40	1406.60
545	1400	1333.80	1390.70	1371.50	590	400	404.31	401.88	396.86
546	1400	1346.10	1402.50	1400.20	591	400	399.95	406.15	410.84
547	1400	1344.60	1392.10	1368.00	592	1400	1356.10	1398.60	1373.40
548	1400	1349.30	1386.80	1374.30	593	400	393.53	395.67	399.35
549	1400	1339.30	1388.90	1370.10	594	1400	1370.80	1411.20	1403.00
550	2500	2467.60	2422.00	2409.70	595	1400	1347.80	1398.30	1374.70
551	1400	1352.40	1403.50	1404.30	596	400	395.60	405.85	411.20
552	1400	1351.60	1354.40	1392.30	597	1400	1391.40	1386.90	1380.90
553	1400	1331.90	1392.80	1359.50	598	400	407.34	401.91	403.42
554	1500	1494.30	1484.70	1483.40	599	1400	1343.80	1396.00	1372.50
555	1500	1485.70	1489.30	1497.70	600	400	402.84	409.91	411.00
556	1400	1350.50	1414.20	1395.20	601	400	399.06	400.01	401.72
557	1400	1351.80	1397.00	1359.30	602	1400	1338.50	1415.40	1373.50
558	2500	2468.40	2465.00	2446.50	603	400	397.49	402.61	400.96
559	1500	1500.90	1491.40	1482.40	604	1400	1350.60	1392.90	1376.60
560	400	402.22	404.77	393.57	605	1400	1336.00	1389.60	1374.50
561	2500	2449.40	2425.90	2409.80	606	1500	1504.80	1519.70	1497.10
562	1500	1527.40	1534.20	1526.60	607	1400	1355.30	1406.00	1391.30
563	1400	1342.40	1391.60	1361.10	608	2500	2542.40	2533.40	2515.90
564	1400	1342.60	1408.60	1394.10	609	1500	1485.50	1484.80	1469.90
565	1400	1353.50	1399.90	1384.10	610	1400	1355.20	1385.20	1379.50
566	1400	1343.30	1389.00	1356.70	611	1400	1355.00	1385.70	1375.80
567	1400	1397.20	1395.20	1387.90	612	1400	1357.00	1382.60	1368.90
568	2500	2462.30	2428.80	2407.80	613	2500	2542.30	2484.00	2464.70
569	1400	1350.70	1404.60	1385.90	614	1400	1352.40	1377.40	1353.80
570	400	391.72	398.02	394.35	615	400	396.15	399.01	400.95
571	400	389.73	395.70	401.23	616	400	390.28	386.80	386.60
572	1400	1330.10	1378.70	1356.60	617	1400	1349.20	1380.20	1379.20
573	400	396.10	404.11	407.63	618	1400	1347.50	1375.30	1358.90
574	1400	1341.30	1385.60	1362.50	619	1400	1345.50	1372.50	1354.20
575	400	397.44	403.45	406.43	620	400	398.33	389.98	392.19
576	1400	1338.00	1396.50	1373.80	621	1400	1353.90	1386.10	1372.60
577	2500	2545.60	2496.10	2469.80	622	1400	1350.40	1380.40	1354.80
578	400	397.75	400.43	403.33	623	2500	2545.10	2494.30	2468.20
579	1400	1363.50	1413.90	1400.90	624	1400	1351.20	1387.10	1368.70
580	2500	2558.40	2515.30	2502.10	625	1400	1353.90	1376.20	1353.00
581	400	400.69	402.09	405.74	626	400	397.15	399.88	402.36
582	1400	1365.10	1410.50	1398.50	627	400	397.23	397.14	396.24
583	1400	1361.70	1403.60	1377.60	628	1400	1352.50	1380.10	1371.90
584	2500	2563.00	2537.30	2527.80	629	1400	1350.20	1383.70	1363.30
585	1400	1367.80	1412.70	1398.30	630	400	394.11	389.10	392.22

Obs	Rec	Z1Thk	Z2Thk	Z3Thk	Obs	Rec	Z1Thk	Z2Thk	Z3Thk
631	400	398.52	396.15	396.90	676	1400	1349.30	1379.90	1383.00
632	1400	1360.10	1385.20	1373.30	677	1400	1349.10	1375.40	1379.10
633	400	399.47	407.33	402.70	678	400	388.83	390.61	397.03
634	1400	1354.60	1387.40	1364.80	679	400	389.27	388.89	396.60
635	1400	1358.80	1382.20	1346.10	680	400	397.99	396.59	402.77
636	1400	1359.00	1387.50	1368.50	681	400	394.91	394.93	403.49
637	1400	1353.30	1382.70	1361.00	682	400	398.94	398.56	406.27
638	2500	2542.50	2491.30	2467.40	683	400	397.21	405.45	404.64
639	1400	1339.70	1384.40	1360.50	684	2500	2417.90	2425.50	2484.20
640	1400	1341.90	1385.00	1357.40	685	1400	1355.10	1418.70	1386.50
641	400	391.80	399.06	393.82	686	400	387.51	395.79	396.01
642	2500	2538.20	2508.70	2488.20	687	400	403.60	409.24	408.59
643	400	399.65	401.32	399.63	688	400	399.37	406.30	403.89
644	1400	1355.00	1381.10	1357.60	689	1400	1364.00	1405.20	1364.40
645	2500	2547.10	2507.90	2484.60	690	400	392.51	393.82	392.50
646	400	404.65	403.40	399.78	691	400	391.96	393.78	396.80
647	1400	1346.90	1387.00	1356.40	692	1400	1388.30	1415.70	1379.80
648	2500	2538.30	2485.40	2461.70	693	2500	2482.10	2515.30	2467.50
649	1400	1358.40	1387.30	1401.20	694	1400	1372.00	1413.80	1389.40
650	2500	2545.60	2515.80	2491.00	695	2500	2462.40	2506.60	2488.40
651	1400	1352.40	1385.50	1366.60	696	1400	1383.30	1401.00	1380.60
652	1400	1352.90	1391.70	1385.10	697	1500	1484.50	1508.80	1528.30
653	400	392.26	392.84	394.90	698	2500	2478.10	2516.30	2463.30
654	1400	1385.80	1365.40	1368.40	699	1500	1488.10	1505.40	1520.80
655	400	392.84	400.46	393.80	700	2500	2484.10	2498.50	2442.60
656	1400	1385.80	1395.40	1367.00	701	1500	1492.90	1512.10	1529.10
657	1400	1346.50	1372.10	1345.30	702	400	390.42	385.72	385.95
658	2500	2500.30	2486.60	2474.90	703	400	390.34	388.14	387.75
659	1500	1516.90	1488.00	1475.30	704	1500	1481.70	1500.50	1509.20
660	2500	2502.40	2474.80	2465.00	705	1400	1386.90	1410.10	1378.60
661	1400	1355.30	1378.40	1362.20	706	400	390.24	390.14	386.20
662	1400	1351.50	1376.30	1351.40	707	1400	1392.80	1415.10	1369.50
663	1400	1359.00	1371.00	1366.00	708	1400	1385.90	1418.70	1389.50
664	1400	1383.00	1407.00	1392.00	709	400	390.47	389.78	388.31
665	2500	2521.70	2509.50	2489.70	710	1400	1388.20	1408.10	1375.30
666	400	393.41	388.70	385.44	711	1400	1391.00	1407.10	1386.30
667	1400	1355.50	1388.20	1355.02	712	1400	1383.60	1410.20	1386.10
668	1400	1357.80	1372.70	1342.90	713	2500	2478.70	2499.00	2439.50
669	1400	1342.70	1383.30	1389.10	714	400	396.48	396.99	396.55
670	1400	1388.00	1391.30	1343.10	715	400	399.98	401.29	398.85
671	1400	1340.20	1371.60	1363.20	716	2500	2470.70	2497.50	2436.30
672	2500	2499.00	2485.70	2502.00	717	400	401.09	403.15	398.07
673	1400	1346.30	1374.50	1360.80	718	400	406.21	408.18	400.86
674	1400	1350.60	1379.10	1381.90	719	1500	1472.10	1496.60	1478.70
675	1400	1386.30	1364.60	1369.30	720	1400	1413.90	1425.80	1408.50

Obs	Rec	Z1Thk	Z2Thk	Z3Thk	Obs	Rec	Z1Thk	Z2Thk	Z3Thk
721	400	411.62	407.14	405.05	766	1400	1383.60	1387.10	1381.30
722	2500	2516.90	2537.60	2516.90	767	400	399.41	393.86	397.57
723	400	402.29	402.52	399.17	768	1400	1369.70	1381.70	1397.80
724	2500	2463.00	2486.90	2457.10	769	400	397.81	393.99	402.61
725	400	401.98	396.11	394.68	770	2500	2471.70	2479.30	2497.50
726	400	397.88	393.72	396.96	771	1400	1387.50	1384.10	1388.20
727	2500	2494.50	2487.60	2439.20	772	1400	1369.10	1371.60	1393.30
728	2500	2492.70	2486.90	2433.10	773	2500	2476.20	2478.50	2494.40
729	400	403.36	401.40	399.20	774	1400	1374.90	1385.00	1393.20
730	1400	1394.50	1406.00	1385.10	775	400	399.03	394.84	401.74
731	2500	2458.00	2479.00	2479.10	776	1400	1383.50	1393.30	1391.50
732	1400	1389.30	1392.90	1364.00	777	1400	1369.10	1386.80	1404.40
733	400	396.27	394.44	392.29	778	2500	2480.40	2487.90	2494.10
734	400	402.64	398.18	396.29	779	400	398.52	393.04	400.06
735	2500	2454.00	2479.00	2455.00	780	1400	1389.10	1399.30	1398.60
736	400	399.14	390.77	387.15	781	400	400.27	393.90	401.51
737	400	399.08	395.02	390.15	782	1400	1394.40	1397.90	1397.70
738	1400	1388.10	1396.80	1355.80	783	400	400.25	394.64	404.21
739	1400	1384.10	1391.00	1353.10	784	1400	1401.30	1365.90	1378.70
740	400	398.03	393.42	392.74	785	400	400.14	396.27	406.79
741	1400	1384.20	1401.80	1374.90	786	2500	2488.90	2479.30	2487.10
742	2500	2463.80	2462.30	2441.30	787	400	400.14	390.84	399.51
743	400	397.23	393.94	392.70	788	400	399.23	392.92	406.76
744	1400	1379.60	1399.00	1396.00	789	2500	2478.90	2480.90	2508.50
745	400	397.70	397.92	395.07	790	2500	2469.00	2480.60	2503.90
746	1400	1385.90	1403.00	1391.80	791	1400	1380.90	1394.60	1410.10
747	1400	1379.90	1385.60	1360.50	792	2500	2492.00	2485.90	2485.50
748	1400	1386.80	1403.90	1391.70	793	1400	1371.20	1382.60	1391.90
749	1400	1373.90	1384.90	1351.30	794	2500	2476.40	2487.00	2512.50
750	400	398.27	400.68	397.83	795	400	406.53	408.46	413.17
751	1400	1387.20	1411.60	1421.70	796	2500	2497.90	2542.30	2531.90
752	400	398.79	395.59	403.79	797	400	404.26	412.08	412.09
753	2500	2484.80	2496.00	2471.60	798	400	404.53	405.71	404.44
754	400	399.49	399.67	392.62	799	2500	2479.30	2523.20	2496.10
755	1400	1369.10	1394.20	1383.60	800	400	408.81	409.37	402.39
756	1400	1378.90	1401.90	1395.10	801	400	409.46	414.69	412.06
757	400	398.92	398.17	397.09	802	400	406.05	411.68	410.31
758	2500	2463.20	2462.60	2418.40	803	400	406.58	414.92	414.78
759	400	400.36	400.52	398.38	804	2500	2500.90	2563.40	2523.10
760	1400	1379.20	1386.50	1360.00	805	400	404.40	407.07	408.72
761	400	402.84	402.54	400.60	806	1400	1396.40	1431.00	1403.70
762	2500	2476.40	2463.80	2434.30	807	1400	1386.60	1428.10	1413.30
763	400	398.54	394.46	396.45	808	400	404.97	408.60	410.57
764	400	402.86	397.49	399.12	809	1400	1391.60	1422.80	1413.90
765	1400	1387.30	1385.80	1370.10	810	400	404.07	403.42	405.80

Obs	Rec	Z1Thk	Z2Thk	Z3Thk	Obs	Rec	Z1Thk	Z2Thk	Z3Thk
811	400	403.60	405.29	410.19	856	1400	1384.90	1387.40	1376.00
812	1400	1398.50	1440.20	1423.60	857	1400	1380.40	1377.80	1366.40
813	400	405.57	404.79	407.70	858	2500	2528.50	2420.40	2409.20
814	400	402.33	401.88	404.25	859	1400	1388.50	1389.00	1366.40
815	2500	2522.90	2560.70	2513.50	860	1400	1391.40	1379.70	1374.60
816	400	401.33	399.58	404.21	861	1400	1428.00	1376.10	1369.50
817	400	404.73	409.82	412.40	862	2500	2532.20	2436.20	2429.70
818	400	405.86	410.52	414.88	863	400	392.80	403.69	406.21
819	2500	2524.70	2480.50	2469.20	864	400	393.91	404.88	407.26
820	400	402.91	404.43	409.29	865	1400	1383.10	1381.70	1390.70
821	1400	1395.70	1432.80	1412.40	866	1400	1397.80	1403.00	1420.70
822	1400	1389.00	1427.30	1425.50	867	2500	2531.20	2521.20	2551.00
823	2500	2516.10	2482.20	2480.70	868	2500	2525.40	2513.20	2549.30
824	400	405.20	412.17	414.91	869	400	412.53	404.81	397.39
825	400	401.00	408.42	411.65	870	2500	2555.50	2515.60	2532.90
826	400	396.54	399.79	406.65	871	1400	1416.40	1395.80	1448.70
827	2500	2487.20	2413.50	2471.50	872	2500	2554.60	2527.30	2541.70
828	1400	1396.00	1396.70	1426.90	873	2500	2494.60	2519.40	2560.20
829	1400	1399.80	1396.00	1402.20	874	1400	1371.10	1395.50	1412.10
830	2500	2535.90	2518.80	2520.30	875	1400	1432.50	1403.40	1424.40
831	1400	1391.50	1387.80	1383.70	876	2500	2477.40	2493.10	2506.10
832	1400	1387.30	1384.40	1363.00	877	1400	1380.40	1395.90	1389.80
833	400	401.60	394.68	400.77	878	2500	2429.60	2423.30	2497.90
834	2500	2525.60	2503.90	2477.10	879	1400	1390.40	1400.80	1435.70
835	1400	1384.80	1393.30	1387.20	880	1400	1376.80	1396.70	1372.90
836	1400	1388.40	1390.50	1388.20	881	1400	1390.80	1397.50	1378.40
837	2500	2522.10	2472.60	2443.00	882	400	390.51	389.35	393.56
838	1400	1388.10	1389.40	1387.60	883	1400	1373.90	1384.60	1365.00
839	1400	1399.90	1389.10	1370.70	884	1400	1375.40	1388.90	1375.90
840	1400	1387.00	1390.00	1382.30	885	2500	2496.90	2422.30	2409.10
841	2500	2491.20	2467.20	2437.50	886	1400	1367.70	1384.00	1367.30
842	1400	1386.60	1380.50	1373.80	887	1400	1377.70	1390.60	1367.20
843	1400	1389.90	1387.00	1371.20	888	1400	1372.40	1393.00	1374.30
844	1400	1389.00	1384.90	1377.90	889	1400	1439.50	1386.80	1371.60
845	2500	2537.30	2429.60	2422.10	890	1400	1370.10	1379.30	1362.90
846	1400	1381.40	1387.60	1385.00	891	1400	1369.80	1382.40	1366.60
847	2500	2520.10	2504.40	2489.40	892	2500	2475.30	2499.80	2460.70
848	400	381.19	381.46	390.36	893	1400	1430.00	1387.70	1373.90
849	1400	1385.20	1382.70	1373.20	894	1400	1359.90	1383.50	1358.80
850	1400	1382.30	1387.70	1381.90					
851	1400	1391.00	1392.50	1382.20					
852	2500	2510.10	2487.20	2476.70					
853	1400	1391.00	1392.50	1382.20					
854	2500	2497.00	2482.80	2464.80					
855	400	386.09	400.90	398.01					

CHAPTER 6

SUMMARY AND CONCLUSIONS

Contributions

This research makes contributions to Multivariate Statistical Process Monitoring in several areas. In Chapter 3:

a) Hawkins' regression adjustment procedure (1993) was generalized to allow its appropriate use for a wider class of cascade processes involving several and varying numbers of measurements at each process step. This procedure involves grouping of the variables according to process step, and performing the regression adjustment across steps.

b) A proof was provided that shows least-squares estimates of the regression coefficients meets conditions for independence between groups being monitored. This enables each chart to be designed so that an overall false alarm rate is maintained.

c) An example demonstrated conditions when the signalling performance of the proposed procedure outperformed traditional monitoring methods and when it did not.

d) The example also demonstrated that for the assumed process model, the proposed method retains diagnostic advantages over T^2 decomposition methods. Diagnostic advantageous over Principal Components Analysis were also demonstrated for cases where principle components contained large factor loadings for variables that spanned more than one group.

In Chapter 4:

a) The novel application (to industrial process data from a local semiconductor manufacturer) of Generalized Linear Models theory to regression adjustment monitoring to deal explicitly with non-normality and non-constant variance in the data was presented. Prediction limits were shown to be tighter than for transformation-based least-squares methods. Though the difference shown in the example was small, it is thought to be a conservative example since departures from normality and constant variance assumptions were small, and models were fit over small range of controlled data. It is hoped that publication of this idea will generate additional interest leading to the analysis of more complex data sets.

b) The example highlighted the appropriateness of prediction interval lengths that vary with the non-constant variance in the original data. Prediction intervals that are inappropriately too tight in regions that naturally exhibit more variability would lead to false alarms. This idea has not been encountered in the literature review as most methods assume normality or make appeals to the Central Limit Theorem. Both transformation-based least-squares and generalized linear models approaches possessed this property.

c) The example demonstrated the flexibility of regression adjustment methods to consider non-linear relationships between quality measurement variables, in contrast to linear-only relationship information in methods that rely primarily on the covariance matrix.

In Chapter 5:

a) The adaptation of regression adjustment for monitoring zone-uniformity when product specifications may change between runs was explored. The method involves adding covariates for product specification to the usual regression adjustment between variables. For the three-zone industrial process considered, loss of uniformity in a single-zone, or in two zones were shown to provide strong signals in the residuals being monitored. A simultaneous shift in all three zones would not be detected, so a T^2 (or other directionally invariant method) running "in parallel" with this approach is recommended.

b) Residuals from the models formed with product specification covariates were shown to provide strong indications of uniformity dispersion effects across product specification levels. In particular, more variability in uniformity was observed with one product specification that was inconsistent with the trend over the other levels -- an indication to process engineers that further investigation may be necessary.

Opportunities for Additional Inquiry

The outcome of any serious research can only be to make two questions grow where only one grew before.

--Thorstein Bunde Veblen (1857-1929)

Multivariate Statistical Process Monitoring of Non-Normally Distributed Quality Characteristics by Monitoring Generalized Variance

It is important to control process variability as well as the mean. When quality characteristics are normally distributed, their means are independent of their variances, and separate monitoring methods are required for each. When the quality characteristics are not normally distributed, their variances are a function of their means. It is reasonable to infer that when non-normally distributed variables are being monitored, an assignable cause would induce changes in both the mean and variance, simultaneously. Furthermore, the covariance between these variables is likely to change in such a scenario -- in other words, the likelihood that the mean and variances of several non-normal variables shift in such a manner that the same covariance structure is maintained is highly unlikely. Since both the variances and covariances are changing, measures that include both could likely be the fastest indication of problem in the process being monitored.

In the multivariate normal case, the sample generalized variance, $|\mathbf{S}|$, which is the determinant of the covariance matrix, is often used to monitor process variability. Montgomery (1991) suggests a control chart based on $E(|\mathbf{S}|) \pm 3\sqrt{V(|\mathbf{S}|)}$, where

$$E(|\mathbf{S}|) = b_1 |\Sigma|, \tag{6-1}$$

$$V(|\mathbf{S}|) = b_2|\Sigma|^2, \tag{6-2}$$

$$b_1 = \frac{1}{(n-1)^p}\prod_{i=1}^{p}(n-i)\text{, and} \tag{6-3}$$

$$b_2 = \frac{1}{(n-1)^{2p}}\prod_{i=1}^{p}(n-i)\left[\prod_{j=1}^{p}(n-j+2)-\prod_{j=1}^{p}(n-j)\right]. \tag{6-4}$$

Over a set of historically "clean" process data, one may also calculate the sample generalized variance, consider these values as "targets," then use procedures involving repeated hypothesis testing that the covariance matrix is equal to that containing the target constants. Details on these procedures are available in Alt (1985) and Morrison (1990).

Reference distributions for control statistics in these methods are based on assumptions that the sample generalized variance is obtained from random samples of p-dimensional multinormal populations. With count, proportion, or otherwise skewed data, this assumption is violated, and robustness to departures have not been studied. Also, under varied quality characteristic distributions, reference distributions for existing test statistics based on the sample generalized variance would be complex (and perhaps impossible to determine), and are likely to be scenario specific. A robustness study and adapted techniques based on empirical reference distributions [similar to Willemain and Runger, 1996] would be interesting and valuable.

Finally, another reason a generalized variance approach may be desireable is when process shift structures are unknown or likely to be in any direction. The method for

dealing with non-normality in Chapter 4 relies on regression adjustment which, in turn, is successful only when usual relationships between variables are violated.

Finite Intersection Test (FIT) for Grouped Regression Adjustment Procedure

Timm(1996) proposed a Finite Intersection Tests (FIT) over the same set of regression adjusted variables proposed by Hawkins (1991, 1993). Timm stated this test was optimal when an a priori order is available (which is the case for cascade processes). As Chapter 3 generalized Hawkins' (1993) procedure, it's reasonable to believe that TIMM's FIT procedure is also generalizable to the case of varying numbers of variables at each process step. The work would most likely encompass adapting the test-statistic shown in (2-29) and determining the correct reference distribution.

Autocorrelation in Regression Adjustment

When measuring the autocorrelation present in Chapter 4, it was noticed that the first-order autocorrelation in the original variables was 0.49, while it was 0.44 in the residuals. Some reduction due to the regression adjustment appears to have taken place, even though only a single regressor was involved. Montgomery and Peck (1992) have noted that autocorrelation in residuals is sometimes an artifact of variable selection. It would be interesting to determine if the amount of residual autocorrelation observed in the quality measurement variables is reduced in the residuals when the number of process steps considered is increased.

Robust Fitting Techniques as a Means of Phase I Retrospective Testing

Initializing statistical process monitoring methods often require use of historical process data to estimate chart parameters. Lack of control in the historical data could cause parameter estimates that result in charts that are insensitive to lack of control in continued monitoring. Montgomery (1991) includes a method for the T^2 that breaks historical data into subgroups for estimating sample means and variance in order to reduce the influence of outliers in the historical data.

In regression, influence of outliers is often reduced by applying a smaller weight to extreme observations, according to some weight function. It would be interesting to investigate Phase I methods that use Robust regression techniques for fitting relationships between the data, then use the weights as a means of identifying "out-of-control" points in the historical data. Estimates of process parameters could then be generated while omitting, or down-weighting these troublesome points.

Independence Between Groups of Residuals in GLM Theory

When using several charts of regression adjusted variables, it is desirable that the residuals in each group are independent of each other, so that each chart may be designed so that an overall false alarm rate may be maintained. This property is present across steps for cascade processes, but is not present when variables at one process step are regressed agains others in the same process step (Hawkins, 1991).

The proof in Chapter 3 demonstrated the presence of this property across groups when least-squares estimates were used. Montgomery and Myers (1997) found some

unexpected results with the use of Generalized Linear Models theory, namely that even though orthogonally designed data was being analyzed, parameter estimates were no longer orthogonal due to the variance functions driving the weights assigned to individual observations during in parameter estimation. It is important to determine if similar impacts on independence between groups of residuals occurs in the Chapter 4 scenario. If the independence property were lost, then observed ARL performance would be different than what was designed.

Conclusion

While several ideas have been presented, they are by no means exhaustive. The earlier discussion of different process structures, shift structures, variable relationships, autocorrelation, data from non-normal distributions with non-constant variance, numbers of variables, etc. create a myriad of opportunities for tailoring procedures to a specific need. The literature review also suggests this field is picking up a "new head of steam."

References

Alt, F. B., (1985). "Multivariate Quality Control," in *Encyclopedia of Statistical Sciences*, Vol. 6, edited by Kotz, S., N.L. Johnson, and C.R. Read, John Wiley, New York, pp. 110 - 122.

Blazek, L. W.; Novic, B.; and Scott, D. M. (1987). "Displaying Multivariate Data Using Polyplots," *Journal of Quality Technology*, 19, pp. 69 - 74.

Brook, D.; and Evans, D. A. (1972). "An Approach to the Probability Distribution of CUSUM Run Length," *Biometrika*, 59, pp. 539-549.

Chih, W. W.; and Rollier, D. A. (1994). "Diagnostic Characteristics for Bivariate Pattern Recognition Scheme in SPC," *International Journal of Quality & Reliability Management*, 11, pp. 53-66.

Chua, M.; and Montgomery, D. C. (1992). "Investigation and Characterization of a Control Scheme for Multivariate Quality Control," *Quality and Reliability Engineering International*, 8, pp. 37-44.

Crosier, R.B. (1988). "Multivariate Generalizations of Cumulative Sum Quality Control Schemes," *Technometrics*, 30, pp. 291-303.

Duncan, A.J. (1986). *Quality Control and Industrial Statistics*, 5th ed. Irwin, Homewood.

Garthwaite, P. H. (1994). "An Interpretation of Partial Least Squares," *Journal of the American Statistical Association*, 89, pp. 122-127.

Geladi, P.; and Kowalski, B. R. (1986). "Partial Least-Squares Regression: A Tutorial," *Analytical Chimica Acta*, 185, pp. 1-17.

Gonzalez, R. C. ; and Woods, R. E. (1992). *Digital Image Processing*, Addison-Wesley, New York.

Harris, T. J.; and Ross, W. H. (1991). "Statistical Process Control Procedures for Correlated Observations," *The Canadian Journal of Chemical Engineering*, 69, pp. 48 - 57.

Hawkins, D. M. (1991). "Multivariate Quality Control Based on Regression-Adjusted Variables," *Technometrics*, 33, pp. 61 - 75.

Hawkins, D. M. (1993). "Regression Adjustment for Variables in Multivariate Quality Control," *Journal of Quality Technology*, 25, pp. 170-182.

Healy, J. D. (1987). "A Note on Multivariate CUSUM Procedures," *Technometrics*, 29, pp. 409 - 412.

Hines, W. W.; and Montgomery, D. C. (1980). *Probability and Statistics in Engineering and Management Science*, Second Edition, John Wiley, New York.

Hotelling, H. (1933), "Analysis of a Complex of Statistical Variables into Principal Components," *Journal of Educational Psychology*, 24, pp. 417-441, 498-520.

Hotelling, H. (1947). "Multivariate Quality Control - Illustrated by the Air Testing of Sample Bombsights," *Techniques of Statistical Analysis*, C. Esenhart, M.W. Hastay, and W.A. Wallis, (eds), McGraw-Hill, New York, pp. 111-184.

Jackson, J. E. (1959). "Quality Control Methods for Several Related Variables," *Technometrics*, 1, pp. 359-377.

Jackson, J. E. (1980). "Principal Components and Factor Analysis: Part I - Principal Components," *Journal of Quality Technology*, 12, pp. 201-213.

Jackson, J. E. (1981). "Principal Components and Factor Analysis: Part II - Additional Topics Related to Principal Components," *Journal of Quality Technology*, 13, pp. 46 - 58.

Jackson, J. E. (1985). "Multivariate Quality Control," *Communications in Statistics - Theory and Methods*, 14, pp. 2657-2688.

Jackson, J. E. (1991). *A User's Guide to Principal Components*, John Wiley & Sons, New York.

Johnson, R. A.; and Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis*, Third Edition, Prentice Hall, New Jersey.

Joliffe, I. T. (1986). *Principal Component Analysis*, Springer-Verlag, New York.

Keats, J. B.; and Shlaes, C. (1994), "A Simple Data Transformation for Use with Statistical Process Control Procedures Applied to Autocorrelated Data," Arizona State University Technical Paper, Department of Industrial and Management Systems Engineering.

Kresta, J. V.; MacGregor, J.F.; and Marlin, T. E. Marlin (1991). "Multivariate Statistical Monitoring of Process Operating Performance," *Canadian Journal of Chemical Engineering*, 69, pp. 35 - 47.

Lowry, C. A.; Woodall, W.H.; Champ, C. W.; and Rigdon, S. E. (1992). "A Multivariate Exponentially Weighted Moving Average Control Chart," *Technometrics*, 34, pp. 46-53.

Lowry, C. A.; and Montgomery, D. C. (1995). "A review of multivariate control charts," *IIE Transactions*, 27, pp. 800-810.

Lucas, J.M.; and Saccucci, M.S. (1990). "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements," *Technometrics*, 32, pp. 1-12.

Mandel, B.J. (1969). "The Regression Control Chart," *Journal of Quality Technology*, 1, pp.1-9.

Mardia, K.V.; Kent, J. T.; and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press, San Francisco.

Mason, R. L.; Tracy, N. D.; and Young, J. C. (1995). "Decomposition of T^2 for Multivariate Control Chart Interpretation," *Journal of Quality Technology*, 27, pp. 99-108.

Mason, R. L.; Tracy, N. D.; and Young, J. C. (1997), "A Practical Approach for Interpreting Multivariate T^2 Control Charts," *Journal of Quality Technology*, In Press.

Mastrangelo, C. M. (1993). "Statistical Process Monitoring of Autocorrelated Data," Dissertation, Arizona State University.

Mastrangelo, C. M.; Runger, G.C.; and Montgomery, D.C. (1996), "Statistical Process Monitoring with Principal Components," *Quality and Reliability Engineering International*, In Press.

Mathcad 3.0, MathSoft, Inc. 201 Broadway, Cambridge, Mass.

McCullagh, P.; and Nelder, J.A. (1989). *Generalized Linear Models*, Second Edition, Chapman and Hall, New York.

Montgomery, D. C. (1991). *Introduction to Statistical Quality Control*, Second Edition, John Wiley, New York.

Montgomery, D.C.; and Mastrangelo, C. M.(1991). "Some Statistical Process Control Methods for Autocorrelated Data [with discussion]," *Journal of Quality Technology*, 23, pp. 179-193.

Montgomery, D. C.; Mastrangelo, C. M.; and Lowry, C. A. (1993). "Statistical Process Monitoring For Dynamic Systems," in *Proceedings of the Institute of Industrial Engineers 2nd Research Conference*, pp. 559 - 563.

Montgomery, D. C.; and Peck, E. A. (1992), *Introduction to Linear Regression Analysis*, John Wiley & Sons, Inc., New York.

Morrison, D.F. (1990). *Multivariate Statistical Methods*, 3rd Edition, McGraw-Hill, New York.

Mortell, R. R.; and Runger, G. C. (1995). "Statistical Process Control of Multiple Stream Processes," *Journal of Quality Technology*, 27, pp. 1 - 12.

Murphy, B. J. (1987). "Selecting Out-of-Control-Variables with T^2 Multivariate Quality Control Procedures," *The Statistician*, 36, pp. 571-583.

Myers, R. H.; and Montgomery, D. C. (1997). "A Tutorial on Generalized Linear Models," *Journal of Quality Technology*, 29, pp. 274-291.

Nelder, J. A.; and Wedderburn, R.W.M. (1972). "Generalized Linear Models," *Journal of the Royal Statistical Society A*, 135, pp. 370-384.

Pierce, D. A.; and Schafer, D. W. (1986). "Residuals in Generalized Linear Models," *Journal of the American Statistical Association*, 81, pp. 977-986.

Pignatiello, J. J. and Runger, G. C. (1990). "Comparisons of Multivariate CUSUM Charts," *Journal of Quality Technology*, 22, pp. 173 - 186.

Pregibon, D. (1980). "Goodness of Link Tests for Generalized Linear Models," *Applied Statistics*, 29, pp. 15-24.

Runger, G. C. (1996a). "Bridging the Islands of Statistical Process Control," Presentation to Quality and Reliability Engineering Seminar at Arizona State University, March 21, 1996.

Runger, G. C. (1996b). "Projections and the U^2 Multivariate Control Chart," *Journal of Quality Technology*, 28, pp. 313 - 319.

Runger, G. C.; and Alt, F. B. (1996). "Contributors To A Multivariate Statistical Process Control Chart Signal, " *Communications in Statistics -- Theory and Methods*, 25, pp. 2203-2213.

Runger, G. C.; and Willemain, T. R. (1995). "Model-Based and Model-Free Control of Autocorrelated Processes," *Journal of Quality Technology*, 27, pp. 283 - 292.

SAS Technical Report P-243 SAS/STAT Software: The GENMOD Procedure, Release 6.09, SAS Institute, Inc., Cary, NC.

Scranton, R.; Runger, G. C.; Keats, J. B.; and Montgomery, D. C. (1996). "Efficient Shift Detection Using Multivariate Exponentially-Weighted Moving Average Control Charts and Principal Components," *Quality and Reliability Engineering International*, 12, pp. 165-171.

Seber, G. A. F. (1984). *Multivariate Observations*, John Wiley and Sons, New York.

Ståhle, L.; and Wold, S. (1988). "Multivariate Data Analysis and Experimental Design in Biomedical Research," *Progress in Medicinal Chemistry*, 25, pp. 291-338.

Stern, H. S. (1996). "Neural Networks in Applied Statistics (with discussion)," *Technometrics*, 38, pp. 205 - 220.

Tiku, M. (1985). "NonCentral Chi-Square Distribution," in *Encyclopedia of Statistical Sciences*, Vol. 6, edited by Kotz, S., N.L. Johnson, and C.R. Read, John Wiley, New York, pp. 276-280.

Timm, N. H. (1996). "Multivariate Quality Control Using Finite Intersection Tests, " *Journal of Quality Technology*, 28, pp. 233-243.

Tong, Y.L. (1990). *The Multivariate Normal Distribution*, Springer-Verlag, New York.

Wade, M. R.; and Woodall, W. H. (1993). "A Review and Analysis of Cause-Selecting Control Charts," *Journal of Quality Technology*, 25, pp. 161-169.

Wang, P.C. (1987). "Residual Plots for Detecting Nonlinearity in Generalized Linear Models," *Technometrics*, 29, pp. 435-438.

Wierda, S.J. (1994). "Multivariate Statistical Process Control - Recent Results and Directions for Future Research," *Statistica Neerlandica*, 48, pp. 147-168.

Willemain, T. R.; and Runger, G. C. (1996). "Designing Control Charts Using an Empirical Reference Distribution," *Journal of Qualtiy Technology*, 28, pp. 31-38.

Williams, D.A. (1987). "Generalized Linear Model Diagnostics Using the Deviance and Single Case Deletions," *Applied Statistics*, 36, pp. 181-191.

Woodall, W.H.; and Ncube, M. M. (1985). "Multivariate CUSUM Quality Control Procedures," *Technometrics*, 27, pp. 285-292.

Zhang, G.X. (1984). "A New Type of Quality Control Charts -- Cause-Selecting Control Charts and a Theory of Diagnosis With Control Charts," *World Quality Congress Transactions*, pp. 175-185.

Zhang, G.X. (1985). "Cause-Selecting Control Charts - A New Type of Quality Control Charts," *The QR Journal*, 12, pp. 221-225.